Modeling Opinion Dynamics on Networks

How Social Influence Shapes the Formation of Consensus and Polarization

– DISSERTATION –

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik, Spezialisierung: *Theoretische Physik*

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
von

M.Sc. Fabian Tilo Werner Baumann

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Igor M. Sokolov
2. Prof. Dr. Francisco A. Rodrigues
3. Dr. habil. Jan Lorenz

Tag der mündlichen Prüfung: 15.06.2021

Abstract

Consensus is key for many aspects of our daily lives. Concepts ranging from language, to social norms, to the rules of a simple card game require population-scale agreement to fulfil their function. People's opinions on major issues of our time should also be reasonably consistent such that important challenges can be tackled collectively. However, this is often not the case. Political opinions as well as the attitudes towards ethical issues are often deeply divided. The origins of such opinion polarization have puzzled sociology for a long time, as social influence often seems to bring us closer together. Hence, what do we have to assume in order to explain opinion cleavages? When will consensus be formed? And what might be the role of increased communication among people in these processes? In this thesis, we tackle such questions by means of mathematical modeling and the analysis of social data. More specifically, we aim to contribute to bridging the gap between microscopic assumptions about social influence and the formation of macroscopic opinion states in social networks. To this end, we first shed light on assimilation processes, where agents' opinions always become more similar upon interactions. Within a formalism based on the spectral decomposition of the underlying network Laplacian, we uncover, and investigate in detail, different mechanisms inhibiting the formation of global consensus states. Typically, however, the assumption of pure opinion assimilation cannot be reconciled with empirical data, where opinions strongly deviate from consensus and are polarized, such that their distribution is highly bimodal and moderate opinions are rare. Therefore, we devise a novel agent-based model to reproduce such stylized facts of controversial debates. The model is based on a minimal set of realistic assumptions: a simple social reinforcement mechanism, homophilic interactions, and a heterogeneous distribution of agents' activities. Contrasted with Twitter data, the model captures empirical features such as bimodal opinion distributions, a clear association between an individual's activity and her conviction, and echo chambers where users are expected to reinforce their opinions due to interactions with like-minded peers. Inspired by the observation of ideological states – where opinions towards different topics are significantly correlated – we extend the model to multiple dimensions. Specifically, we assume that agents' opinions evolve in a multi-dimensional opinion space which is spanned by base vectors representing the topics discussed. Within this framework we can qualitatively reproduce some opinion distributions observed in the 2016 American National Election Survey, including ideological states that emerge in the model when the orthogonality assumption of topics is relaxed. On the theoretical sides, our work sheds light on some analytically tractable models of networked dynamical systems. More

practically, through the links between simple and interpretable models of opinion dynamics and empirical social data, we aim to contribute to a deeper understanding of collective opinion formation in society.

Zusammenfassung

Konsens ist entscheidend für viele Aspekte unseres täglichen Lebens. Von der Sprache über soziale Normen bis hin zu den Regeln eines einfachen Kartenspiels ist breite Übereinstimmung erforderlich, um den jeweiligen Zweck zu erfüllen. Auch die Meinungen zu den zentralen Themen unserer Zeit sollten so weit übereinstimmen, dass wichtige Herausforderungen gemeinsam in Angriff genommen werden können. Dies ist jedoch häufig nicht der Fall. Sowohl politische Meinungen als auch Einstellungen zu ethischen Fragen sind oft tief gespalten. Die Ursprünge einer solchen Meinungspolarisierung haben die Soziologie lange Zeit vor ein Rätsel gestellt, da soziale Einflüsse uns oft näher zusammenzubringen scheinen. Was müssen wir also annehmen, um Meinungsspaltungen zu erklären? Wann wird ein Konsens gebildet? Und welche Rolle könnte die gestiegene Kommunikation zwischen Menschen bei diesen Prozessen spielen? In dieser Arbeit werden wir uns solchen Fragen mit Hilfe mathematischer Modellierung und der Analyse sozialer Daten nähern. Im Speziellen soll dazu beigetragen werden, die Lücke zwischen mikroskopischen Annahmen über den sozialen Einfluss und der Bildung von makroskopischen Meinungszuständen in sozialen Netzwerken zu schließen. Zu diesem Zweck beleuchten wir zunächst Assimilationsprozesse, bei denen sich die Meinungen von Agenten durch Interaktionen immer weiter angleichen. Innerhalb eines Formalismus, der auf der spektralen Zerlegung der Laplace-Matrix des Netzwerks basiert, decken wir verschiedene Mechanismen auf, die die Bildung von globalen Konsenszuständen verhindern und untersuchen diese im Detail. Typischerweise lässt sich die Annahme einer reinen Meinungsassimilation jedoch nicht mit empirischen Daten vereinbaren, bei denen Meinungen stark vom Konsens abweichen und polarisiert sind, ihre Verteilung also stark bimodal ist und moderate Meinungen selten sind. Aus diesem Grund entwickeln wir ein neuartiges agentenbasiertes Modell, um einige stilisierte Fakten kontroverser Debatten zu reproduzieren. Das Modell basiert auf einem minimalen Satz realistischer Annahmen: einem einfachen sozialen Verstärkungsmechanismus, homophilen Interaktionen und einer breiten Aktivitätsverteilung der Agenten. Verglichen mit Twitter-Daten, erfasst das Modell empirische Merkmale wie bimodale Meinungsverteilungen, einen klaren Zusammenhang zwischen der Aktivität eines Individuums und ihrer Überzeugung sowie Echokammern, in denen Meinungen durch die Interaktionen mit Gleichgesinnten verstärkt werden. Inspiriert durch die Beobachtung von ideologischen Zuständen – wo Meinungen zu verschiedenen Themen signifikant korreliert sind – erweitern wir das Modell auf mehrere Dimensionen. Insbesondere gehen wir davon aus, dass sich die Meinungen der Agenten in einem mehrdimensionalen Meinungsraum entwickeln, der durch

Basisvektoren aufgespannt wird, die die diskutierten Themen repräsentieren. Innerhalb dieses erweiterten Modells können wir Meinungsverteilungen, die wir im American National Election Survey aus dem Jahr 2016 beobachten, qualitativ reproduzieren, einschließlich ideologischer Zustände, die im Modell dann auftreten, wenn Themen nicht mehr orthogonal sein müssen. Aus theoretischer Sicht beleuchtet diese Arbeit einige analytisch behandelbare Modelle vernetzter dynamischer Systeme. Durch die Verknüpfung von einfachen und interpretierbaren Modellen der Meinungsdynamik und empirischen sozialen Daten, möchten wir zusätzlich zu einem tieferen Verständnis der kollektiven Meinungsbildung in der Gesellschaft beitragen.

Selbstständigkeitserklärung

Ich erkläre, dass ich die Dissertation selbständig und nur unter Verwendung der von mir gemäß § 7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 42/2018 am 11.07.2018 angegebenen Hilfsmittel angefertigt habe.

Berlin, 24. August 2021 Fabian Baumann

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Igor Sokolov, who always had an open door to discuss. He gave me a lot of freedom to follow my interests and was an important help to put our ideas on a firm footing. Speaking of ideas; these would not have been developed, let alone ended up in two papers, without the incredible drive and enthusiasm of my old friend and colleague Philipp Lorenz-Spreen. A big thank you also goes to Michele Starnini, who believed in our work from the very start and had the necessary persistence to go the extra mile it sometimes takes. It was also a great joy to work with Melvyn Tyloo, and I really hope we will get the chance to collaborate in the future. Unfortunately, my time in Brazil was interrupted by Corona. I would have very much enjoyed to continue working with Prof. Francisco Rodrigues and Thomas Peron, who made me feel very welcome during my short stay in São Carlos. For the corrections of many typos and valuable suggestions on this thesis and their support I would like to thank (again) Michele, Philipp, Adrian, Melvyn, Andreas, and Joni.

Finally, I would like to remind my family, and especially my mum, for being my great motivation to keep going and I thank them from the bottom of my heart for their unconditional support. Most importantly, thank you Lea for that you have always been there despite all my weird moods – hopefully one day you can forgive me for not realizing what's really important so we can laugh together again.

Berlin, August 24, 2021 Fabian

This work was supported through the DFG by the IRTG 1740.

Contents

List of Symbols and Abbreviations

ABM	agent-based model
VM	voter model
MRM	majority rule model
CODA	continuous opinions and discrete actions
PAT	persuasive arguments theory
ER	Erdős-Rényi (model/network)
BA	Barabási-Albert (model/network)
WS	Watts-Strogatz (model/network)
SBM	stochastic block model
AD	activity-driven (model/network)
MRD	modified resistance distance
DO	damped oscillator model
SOC	second-order consensus model
ANES	American national election survey
N	system size (number of nodes or agents)
t	time
$x_i(t)$	opinion (or state) of agent i at time t, often simply x_i
$\mathrm{sgn}(x_i)$	opinion stance of agent i
$\lvert x_i \rvert$	opinion strength, or conviction, of agent i
x_i^∞	final opinion of agent i
$\mathbf{x}(t)$	opinion (state) vector at time t
$\mathscr{G}$	graph
$\mathscr{N}$	set of nodes
N_E	number of edges in a network
$\mathscr{E}$	set of edges
$\mathbf{A}$	adjacency matrix
$\mathbf{A}(t)$	time-varying adjacency matrix
A_{ij}	ijth element of the adjacency matrix
δ_{ij}	Kronecker delta
k_i	degree of node i
$p(k)$	degree distribution
$\mathbf{W}$	weighted coupling matrix
$\mathbf{W}(t)$	time-varying coupling matrix
s_i	strength of node i
$\mathbb{L}$	network Laplacian
$\mathbb{L}_\mathrm{w}$	weighted network Laplacian
$\mathbb{L}^\dagger$	Moore-Penrose pseudo inverse of $\mathbb{L}$
$\mathbf{D}$	diagonal degree matrix
λ_α	αth eigenvalue of the network Laplacian
$\mathbf{u}_\alpha$	αth eigenvector of the network Laplacian
$c_\alpha(t)$	αth (time-dependent) expansion coefficient of $\mathbf{x}(t)$ over $\{\mathbf{u}_\alpha\}$
Ω_{ij}	resistance distance between two nodes i and j

Γ_{ij}	path from node i to node j
l_{ij}	shortest path from node i to node j
c_i	local clustering coefficient of node i
Q	network modularity
p_{ER}	connection probability between two nodes (ER model)
p_{ws}	rewiring probability (Watts-Strogatz model)
m_{BA}	number of links generated by a newly added node (BA model)
b	number of blocks in the (SBM)
p_{intra}	connection probability between two nodes within one block for $b = 2$ (SBM)
p_{inter}	connection probability between two nodes in two different blocks for $b = 2$ (SBM)
a_i	activity of agent i (AD model)
$F(a)$	activity distribution (AD model)
m	number of links an active agents generates (AD model)
V_s	set of stubborn agents
N_s	number of stubborn agents
$P_i(t)$	(time-dependent) bias opinion of stubborn agent i
$\mathbf{P}(t)$	(time-dependent) vector of opinion biases, often simply $\mathbf{P}$
κ	stubbornness of biased agents
K	diagonal stubbornness matrix
$\mathbb{L}^{(\kappa)}$	modified Laplacian
$\lambda_\alpha^{(\kappa)}$	αth eigenvalue of the modified Laplacian
$\mathbf{u}_\alpha^{(\kappa)}$	αth eigenvector of the modified Laplacian
$\Omega_{ij}^{(\kappa,p)}$	pth-order modified resistance distance between node i and j
$\rho(\Omega_{ij}^{(\kappa,1)})$	probability density of 1st order MRDs
$C_p(i,V_s)$	pth-order centrality of node i based on MRDs
$\Theta(x)$	Heaviside step function
$\mathscr{C}$	coherence measure of consensus change
$\tilde{\mathscr{C}}$	$\mathscr{C}$ averaged over all nodes in the network
$\langle\tilde{\mathscr{C}}\rangle$	$\tilde{\mathscr{C}}$ averaged over multiple network realizations
$D_{\max}$	maximum opinion distance
$\tilde{D}_{\max}$	$D_{\max}$ averaged over all possible pairs of nodes
$\langle\tilde{D}_{\max}\rangle$	$\tilde{D}_{\max}$ averaged over multiple network realizations
μ_x	final mean opinion
$\tilde{\mu}_x$	μ_x averaged over all possible pairs of nodes
$\langle\tilde{\mu}_x\rangle$	$\tilde{\mu}_x$ averaged over multiple network realizations
σ_x^2	final opinion variance
$\tilde{\sigma^2}_x$	σ_x^2 averaged over all possible pairs of nodes
$\langle\tilde{\sigma^2}_x\rangle$	$\tilde{\sigma^2}_x$ averaged over multiple network realizations
n_r	number of randomly rewired edges
$f(t)$	time-periodic modulation function of the coupling weights
ω	modulation frequency
ω_α^*	resonant modulation frequency associated with αth mode
h	modulation amplitude
γ	coupling strength of agents' positions
μ	coupling strength of agents' velocities
d	damping coefficient
$\mathbb{L}(t)$	time-dependent network Laplacian
$\mathbb{L}(0)$	static network Laplacian (equivalent to $\mathbb{L}$) defined for clarity in distinction to $\mathbb{L}(t)$
$\lambda_\alpha^{(0)}$	αth eigenvalue of $\mathbb{L}^{(0)}$

$\lambda_\alpha(t)$	time-dependent αth eigenvalue of $\mathbb{L}(t)$
$k(t)$	time-dependent damping coefficient
q_α	transformation of c_α
$K(t)$	transformation of $k(t)$
$\Omega_\alpha(t)$	time-dependent eigenfrequency of Hill equation for q_α
ε	detuning from resonance frequency
s_α	amplitude growth exponent for resonant modulation
$\mathscr{A}(\omega)$	integrated amplitude response at ω
$\langle\cdot\rangle_{\{IC\}}$	average over multiple initial conditions
$\mathscr{I}_{\text{init}}$	interval from which initial conditions are sampled
R	Kuramoto order parameter
$\varphi(t)$	time-dependent phase of oscillator i (Kuramoto model)
d_c	confidence bound
K	social interaction strength
α	controversialness of topic/discussion
γ	power law exponent of the activity distribution
ε	lower bound of the power law activity distribution
p_{ij}	connection probability from agent i to j (homophily)
β	homophily exponent
r	reciprocity
$\langle a\rangle$	average activity
α_c	critical controversialness
$\lvert\langle x_f\rangle\rvert$	absolute value of the average final opinion
$\mathbb{J}$	Jacobian of coupled dynamical system
$\tilde{\lambda}$	largest eigenvalue of the Jacobian
$P(x)$	opinion distribution
$\langle x\rangle^{NN}$	average opinion of nearest neighbors of an agent
A_{ij}^{agg}	aggregated interaction network
$\mathbf{x}_i(t)$	time-dependent opinion vector of agent i containing opinions towards different topics
x_i^v	opinion of agent i towards topic v
$\mathscr{T}$	(multi-dimensional) topic space
T	number of topic dimensions
Φ	topic overlap matrix
δ_{uv}	topic angle between topics u and v
$\cos(\delta_{uv})$	overlap between topics u and v ($\Phi_{uv} = \cos(\delta_{uv})$)
$d(\mathbf{x}_i, \mathbf{x}_j)$	opinion distance between agent i and j
$P_u(x)$	opinion distribution with respect to topic u
$\sigma_u^2(x)$	variance of opinions with respect to topic u
$\rho(x^{(u)}, x^{(v)})$	opinion correlation between opinions towards topics u and v
r, φ	polar opinion coordinates

List of Figures

Introduction

Social influence is everywhere. It acts in virtually any encounter humans have with each other and shapes our opinions on important political, ethical, and societal issues, both on the level of individuals, but also more macroscopically in larger groups or society as a whole [1, 2, 3, 4]. Since the beginning of the 21st century omnipresent digitalization has fundamentally changed the way we interact and potentially amplified the importance of social influence. While people communicate with ever increasing volumes, both at home and on the move, and across geographical barriers [5, 6, 7], this communication has also novel properties. In particular, online social media platforms, such as Facebook, Twitter, and Instagram empower individuals to create and disseminate new content, which is consumed by others at unprecedented rates [8]. Moreover, single users on such platforms may have an enormous outreach compared to the offline world, facilitating the spread of content and potentially influencing a very large audience [9]. These developments did not only revolutionize the advertising of commercial products through influencers [10], but also substantially transformed our information landscape with important implications for public opinion formation. Traditional news media, such as newspapers, radio stations, and linear television, which have acted as gatekeepers and crucial filters of information flows [11], were replaced, or at least complemented, by decentralized (online) information sources and individuals promoting certain viewpoints [12]. Besides several positive effects, such as helping to establish peoples' movements against autocratic regimes as "liberation technology" or the democratization of information [13], social media has also contributed to the spread of biased information and the dissemination of conspiracy theories or false news [14, 15, 16]. Importantly, these characteristics of modern communication infrastructures may turn public opinion formation into an increasingly self-organized process, which is substantially driven by the social interactions among individuals.

Opinions on different important issues are not just internal cognitive concepts. They may lead to individual and collective action with major consequences for the stability and the functioning of democracy and society. According to Georg Simmel, a 19th century born sociologist, societies "need some quantitative ratio of harmony and disharmony" in order to persist [17, 18]; a quote, which may be understood as follows. On the one hand, "harmony", also referred to as *consensus* in the following, is indeed important. It includes the agreement on various cultural norms, conventions, and shared beliefs, to ensure the functioning of society as a whole [19, 20]. More specifically, consensus on pressing issues is necessary, since many global challenges, such as the climate change [21], the loss of biodiversity [22],

or the outbreak of global pandemics [23], require the collective action and can only be met as an international cooperative in a coordinated way [24]. On the other hand, too much homogeneity and agreement can be disadvantageous. For example, a lack of plurality with respect to different political choices would declare democracy in itself to be meaningless [25]. It was furthermore argued that a certain degree of dissent may promote the positive aspects of a liberal democracy, where grievances are resolved through constructive debates and the exchange of opposing arguments [26]. Lately, however, there has been very limited concern about *too much* societal consensus and agreement; quite the contrary. The U.S. and Europe have witnessed increasingly diverging viewpoints on various political issues. Such strong forms of opinion polarization have been characterized as predominantly harmful to society; its anticipated effects range from a decrease in the quality of policy choices made by decision makers, to political gridlock [27], to the collapse of democracy [28]. By that, any potentially positive effects of moderate levels of societal dissent are vitiated [26]. Increasing opinion polarization has also been linked to political radicalization [29], which may have contributed to recent disturbing events such as the storming of the German Reichstag in August 2020 and of the U.S. Capitol in Washington D.C. by right-wing extremist groups several months later.

Traditionally, research on opinion formation has been situated within the social sciences, where it is typically performed at two different scales. On the one hand, survey-based research helps to uncover public attitudes towards different political or ethical issues [18, 27]. By considering representative samples of individuals one obtains an aggregated view on *society as a whole*. This macroscopic approach is mostly taken by sociologists and political scientists and puts only very limited emphasis on individuals and their mutual interactions. More specifically, instead of considering the opinion of a single individual, the focus is on the distribution of opinions towards a specific topic. However, every society is made up of individuals which interact and socially influence each other. Thus, the question is: How do single individuals react to actions, or the expression of their peers' opinions? Questions as the latter one have been tackled by social psychologists, identifying social influence between individuals as an important driver for opinion formation. A whole zoo of social influence mechanisms has been established, including findings on persuasive-arguments theory [30, 31], imitation mechanisms and social learning [32], group pressure [33], leader-follower effects [34], and cognitive consistency theories [35, 36]. Social psychological research is typically based on experiments of dyadic social influence (involving only two individuals) or small groups of individuals. Such well-controlled lab experiments have one very important advantage: they allow for strong causal inferences with respect to social influence mechanisms [4]. However, it is not possible to obtain insights into how social influence on the micro-scale determines macroscopic opinion states involving many individuals, or how those states would develop over time [3, 4].

How can physics contribute to the investigation of social systems, and, more specifically, to the understanding of consensus formation and the emergence of opinion polarization in modern societies? The short answer is "mathematical modeling of complex systems". The more extensive answer discusses the role of physics in the interdisciplinary effort to bridge

the *micro-macro gap* within the social sciences between (i) social influence mechanisms on the micro scale (informed by social psychology), and (ii) collective, or aggregated, social phenomena on the macro-scale [4, 37]. This mechanistic approach is based on the assumption that the macroscopic state of a social system is not determined by a centralized authority controlling its dynamics. Instead, social macro-states are assumed to emerge in a self-organized manner from repeated interactions among individuals [3, 4, 20, 38]. This paradigm has deep connections to more traditional working grounds of physicists. Since the 19th century, statistical physics has been developing mathematical frameworks to connect the micro- to the macro-scale in order to understand the behavior of different complex systems [39, 40]. While the focus was first on purely physical problems, such as the development of a kinetic theory of gases [41], statistical physics has ever since been applied to highly contemporary fields ranging from neuroscience [42] to evolutionary biology [43]. Inspired by philosophers, such as David Hume or Auguste Comte, who speculated about a new kind of science (the term "sociology" was coined by Comte) according to which society is governed by a general set of laws, which may be formulated in the language of mathematics – *la physique sociale* [44, 45] – physicists did not hesitate long to apply their powerful mathematical tool sets also to questions in the social sciences.

Early works of socio-physics starting from the middle of the 20th century mainly aimed for explanations on how order, or other global patterns, may arise in social systems from local interactions [46]. Prominent examples range from the emergence of languages [47], pedestrian flows [48], cultures and hierarchies [38], to the formation of consensus in opinion dynamics [49, 50], which were mainly approached with models previously used in a statistical physics context, giving rise to disorder-order phase transitions. To apply such models to social systems rather crude and abstract analogies were employed [38, 51]. For instance, the voter model [49] and the majority rule model [50] were used to model opinion dynamics; they are both based on the Ising model originally introduced to explain ferromagnetism [52]. Inspired by the discrete nature and the simple spin interaction dynamics of the Ising model, it is assumed that binary opinions, which are placed on a grid, interact via (anti-)ferromagnetic interactions. This formalism models situations in which individuals aim to become more (dis)similar to their connected peers. While those models could be analysed using tools from statistical physics, they have also been strongly criticized for their lack of transferability to actual sociological setups, and it was argued that more realistic, and empirically grounded, assumptions about social systems need to be included [51, 53]. An important limitation of early physics-inspired models concerns the discreteness of the underlying opinion space, which precludes the implementation of realistic features of social influence between individuals. A binary opinion as in the voter model does not have a strength and is simply represented by its sign. Thus it may only switch back an forth between two completely antagonistic opinions. By contrast, a more realistic model could include individuals which might go through phases of confidence and indecision [54], or may radicalize as their convictions become more extreme. Most importantly, however, early opinion dynamics models suffered most from a lack of empirical data, allowing to contrast the model with stylized facts of real opinion formation

processes.

These latter drawbacks are starting to be addressed as the data revolution, which has unfolded since the beginning of the 21st century, offers unprecedented opportunities for the investigation of social systems outside the laboratory. It is undisputed that individuals and their interactions are incredibly complex. This is due to the very large number of variables and important features of social dynamics that are often unknown and hard to measure [55]. Furthermore, social systems do generally not come in well-controllable ensembles, which allow for repeated and independent measurements. Moreover, the behavior and interactions of humans are influenced by individual differences and various "hidden" cognitive processes and emotions, which are time-dependent and generally very hard to quantify in experiments [56]. Nevertheless, the massive data sets on large populations often give rise to predictable collective behaviors and social pattern [38]. Striking regularities have been found in social systems ranging from Zipf's law in quantitative linguistics [57], to the scales of human mobility [58], to power law distributions of activity in social interactions [59]. This *mining of the social reality* of a large number of individuals, initiated the rise of the new interdisciplinary research field of computational social science, which complements "traditional" modeling approaches of social systems, by data-driven analyses [60]. For instance, GPS data of mobile phones and "smart" badges have offered new possibilities to study the movements of people as well as their face-to-face interactions. Besides the analysis of human movement pattern in physical space, digital technologies have proven valuable for the collection of macro communication pattern involving data of phone calls or email communication [60].

A natural representation of data sets containing information on complex social systems is offered by network science [61, 62]. It's idea is to model social data as networks, with nodes typically representing individual humans and edges encoding their relations and interactions [63]. With its origin in sociograms developed by Jacob Moreno in the 1930s [64], social network analysis has become a versatile tool to study interconnected social systems. Traditional social network studies were often based on, and complicated by, tedious and error-prone data collection procedures [65]. They involved surveys and in-person interviews limiting social network analyses to rather small systems. Due to enhanced computing capabilities and the overabundance of digital social data also the analysis of large-scale social networks became feasible [61, 62]. Prominent examples range from networks of collaboration of movie actors [66, 67], coauthorship networks among academics [68, 69], or the mapping of the world wide web [70], which has, besides its technological character, also a social component to it. More recently, social networks reconstructed from data provided by social media platforms, such as re-tweet [6] and follower-followee networks (Twitter) [71], or friendship networks (Facebook) [16, 72], have offered new possibilities to investigate social dynamics such as the spreading of (mis-)information and collective opinion formation [16, 72]. In addition to exploring the structural properties of many social networks, it was soon realized that those properties, such as very short average path lengths [66], power law degree distributions [73, 74], or pronounced community structures [75] profoundly impact the dynamical processes unfolding on top of them, including the formation of opinions [38, 76, 77, 78, 79, 80].

The research presented in this thesis is situated between theoretical contributions to established models of opinion dynamics, network science and more data-driven approaches in computational social science. We contribute to the understanding of models of opinion dynamics which aim to overcome some crucial limitations of early approaches. First, more realistic models should take into account structural features of real social networks, which have been found to show strong deviations from simple graph structures, i.e. regular lattices or well-mixed populations, as often considered in early physics-inspired models of social systems. Secondly, empirically grounded social influence mechanisms, informed by cognitive science, should be considered for modeling the evolution of individual agents' opinions as well as their mutual interactions [4, 51]. One of these processes is *opinion assimilation*, where individuals' opinions become more similar upon interaction. Pioneered by French [81], Harary [82], and DeGroot [83], such models have mostly been implemented as discrete-time *opinion averaging* on networks. If time is viewed as a continuous variable, such assimilation models give rise to a diffusive coupling between agents' opinions which is induced by the network Laplacian. Within a spectral decomposition approach, we study different continuous-time assimilation models analytically and contribute to the understanding of the interplay between network structure and the formation of different collective opinion states.

Opinion assimilation models, however, do not tell the whole story. As formally derived by Abelson, those "classical" models *always* reach a global consensus if the underlying social influence network is connected. This led Abelson to wonder "*(...) what on earth one must assume in order to generate the bimodal outcome of community cleavage studies*" [84]. Previous empirical studies, and also our own analyses of Twitter data and data of the 2016 American National Election Studies (ANES), show that opinion distributions often deviate from a global consensus. Instead, opinions are distributed in a highly heterogeneous manner and distributions are often strongly polarized, especially on controversial issues. Also other phenomena are observed which typically cannot be reconciled with opinion assimilation, including the association between the activities and the opinions of individuals [72, 85], or echo chambers [85, 86, 87, 88, 89], where the opinions of individuals are reflected in the opinions of their peers in the social network. It can be furthermore observed that the opinions towards different topics are sometimes strongly correlated. To understand such phenomena from a mechanistic perspective, we propose a simple and interpretable model based on minimal but realistic assumptions on social influence and the interaction dynamics between individuals. In particular, the model is based on a simple social reinforcement mechanism and interactions which are driven by the activity of agents and ruled by homophily.

Structure of the thesis. In Ch. 1, we give a brief introduction to agent-based modeling in opinion dynamics and introduce some important concept of social psychology. In Ch. 2, we first review the basic definitions necessary to describe networks algebraically in terms of matrices. Subsequently, we discuss crucial properties of social networks as well as some generative models to reproduce their most important features. These two chapters serve as a general introduction to opinion dynamics and network theory and can therefore be skipped by readers familiar to these subjects.

Part I focuses entirely on linear assimilation models which are introduced from an opinion dynamics perspective (Ch. 3). Thereafter, in Ch. 4, we present analytical results on a classical opinion assimilation model focusing on the formation of non-consensus states induced by stubborn agents. Chapter 5 represents an exception as it does not deal with opinion dynamics. Instead, we consider consensus models, which have previously been applied in technical applications such as multi-vehicle formation control. Using a general second-order consensus model, we uncover a mechanism that inhibits the formation of consensus for small time-periodic modulations of the coupling strength.

In Part II, we consider non-linear models of opinion dynamics and aim to follow a more data-driven approach, where model outcomes are compared to empirical opinion data. First, we review some relevant (non-linear) extensions to classical assimilation models (Ch. 6). Subsequently, we introduce a novel model based on homophily and a simple social reinforcement mechanism. In Ch. 7, the model is discussed in one dimension, where it is able to qualitatively reproduce important features of polarized debates on Twitter. In Ch. 8, motivated by the assumption that topics are rarely discussed in isolation, the model is generalized to opinion spaces of multiple dimensions. We find that this natural generalization may gives rise to ideological states where opinions towards different topics are highly correlated. The analysis of questions from the 2016 American National Election Survey, reveals similar states, by which we shed light on the formation of opinion correlations in debates around controversial issues. In Ch. 9, we conclude the work and summarize our main findings. As an outlook, we give some perspectives on how our results may relate to future developments in the field of computational social science.

1 From Social Influence to Opinion Dynamics

There is no universally accepted definition for the concept of an opinion. We will therefore use the term "opinion" in a rather general manner. For instance, our opinions may determine how we would answer questions like "which computer operating system is better, Windows or Linux?" [38]; "how do you think about abortion"; "is voting a duty or a choice?"; "how much should the state invest in renewable energies?"; or "should transgender people (...) be allowed to use the bathrooms of their identified gender?" [1]. While opinions sometimes have a strong binary character ("Windows"/"Linux", "choice"/"duty", yes/no, or pro/con), they often contain a strength that we will also call conviction. For example, some people very strongly believe that voting is a duty, others, by contrast, have a rather moderate opinion on this issue, or are even undecided. Generally, a person's opinion summarizes their beliefs, judgements, and emotions towards a certain issue or topic, which is mostly uncertain and subjective and can therefore not be proven [91].

Different disciplines, ranging from political science to social psychology have investigated how the opinions of individuals form. Certainly, there are various factors contributing to this process. While social psychologists have mostly focused on the microscopic, or dyadic, interactions between two individuals mediating social influence [4], political scientists have taken a more macroscopic approach emphasizing the effects of e.g. mass media or the role of influential politicians [92, 93]. Independently of any social process, it has also been pointed out that genetic factors may play a role in determining individuals' opinions on some political issues [94].

The general idea of models of opinion dynamics is a different one. Here the process of individual and collective opinion formation is shaped by the social interactions within large groups of individuals [38, 55]. During their interactions, individuals socially influence each other and thereby update their opinions. Inspired by empirical observations, different social influence mechanisms have been considered. For instance, it may be assumed that individuals align their opinions to become more similar, or reinforce them upon interacting with similar peers [3, 4]. Social influence is not necessarily symmetric, i.e., there may be cases, where a

[1] These latter examples are taken from the 2016 American Nation Election Studies [90], which we analyze in Ch. 8.

pairwise interaction induces an opinion change only in one of the two individuals, or the opinion of one individual changes more strongly than their peer's opinion [4].

Agent-based modeling

Having defined appropriate microscopic interaction rules, i.e., the social influence mechanisms, mathematical models allow us to explore how macroscopic social phenomena emerge. Often, one formulates such approaches as *agent-based models* (ABM), where each "real" individual is represented by an agent in the model [95]. Outside of opinion dynamics, ABMs have been used to investigate a wide range of social and economic phenomena including social segregation [96], cultural differentiation [97], traffic and pedestrian dynamics [55, 98, 99], stock market trading [100], or the diffusion of innovation [101] to name just a few. In all those domains, ABMs have served as vital tools for empirical research towards "generative" social science [37, 102, 103], which defines a promising bottom-up approach, whose overarching goal is concisely summarized by the title of T.C. Schelling's book "Micromotives and macrobehavior" [104]. First, we define microscopic properties of individual agents *and* their interactions. Then, the system evolves according to those rules and gives rise to emerging macroscopic phenomena.

Generally, an agent-based model considers ensembles of N agents, where each agent $i = 1,\ldots,N$ is potentially characterized by different properties. For instance, agents may have different levels of activities to engage in social interactions [105, 106, 107], different spatial positions or group memberships as in models of social segregation [96], or different cultural traits as in Axelrod's model of cultural dissemination [97]. In opinion dynamics, the focus is on the variable $x_i(t)$. It defines the time-dependent opinion of agent i towards one issue or topic. In this thesis, we assume that the change of agent i's opinion due to agent j can be described by a function which depends solely on the opinions x_i and x_j. In contrast to some higher-order influence models [108], this assumption excludes cases where the social influence between two agents i and j depends on the opinions of further agents k that interact simultaneously with agents i and j. Traditionally, most studies on opinion dynamics have considered one-dimensional opinions. However, more recently several multi-dimensional models have been proposed [109, 110, 111, 112, 113, 114, 115]. Here, the idea is that agents do not only hold a single opinion as in one-dimensional models, but multiple opinions on different issues or topics. In this thesis, we will first consider one-dimensional models of opinion dynamics. In Ch. 8, we develop a simple multi-dimensional model in order to explain the emergence ideological opinion states, where agents' opinions towards multiple topics are correlated [116, 117, 118].

Models of opinion dynamics can be well categorized in terms of the underlying opinion space. Different types of opinion spaces have been considered. One model class assumes that agents hold discrete opinions, which are usually drawn from finite sets. The most well-known representatives of this kind are the voter model (VM) [49, 119] and the majority rule

model (MRM) [50], where agents' opinions are binary, i.e., $x_i(t) \in \{-1, 1\}$. In the voter model, opinions evolve as follows: in each time step, a randomly selected agent adopts the opinion of a randomly chosen nearest neighbor. This update rule is similar to the dynamics of the Ising model [52], where – in the ferromagnetic case – aligned spins are energetically favored. Note that, usually no bulk noise is considered for the VM, hence, consensus states, with $x_i = 1$ or $x_i = -1 \forall i$, are absorbing [38]. In the MRM, we select a random agent together with their nearest neighbors on a network, or we choose random subsets of agents in a well-mixed population. All chosen agents update their opinion simultaneously, such that each agent adopts the opinion of the prevailing majority in the corresponding group. Discrete opinion spaces restrict the set of possible social influence mechanisms. Specifically, in the binary case opinions can only align or anti-align due to positive and negative social influence, respectively.

Continuous models of opinion dynamics with $x_i(t) \in \mathbb{R}$ are more flexible. They allow to implement various more refined mechanisms, such as opinion assimilation. Here, one assumes that the opinions of agents gradually approach each other. Continuous opinion variables also allow us to implement similarity biases, where agents only interact if they have sufficiently similar opinions, as we will discuss in Ch. 6. Another important mechanism, which cannot be implemented in discrete models of opinion dynamics is group polarization [120]. Here single individuals develop more extreme opinions after interacting with similar peers. This process is at the core of the radicalization model introduced in Ch. 7. Continuous opinion scales have further important advantages for contrasting modeling results to empirical research on opinion changes [4]. Typically, in social psychological experiments one considers individual opinions towards a certain issue or topic, which may vary on a continuous scale. For instance, they may range on a 1-to-9 scale, where 5 defines a neutral opinion, and the values 1 and 4 correspond to strongly and mildly against the issue, respectively. Likewise, values larger than 5 define opinions in favor of the issue [4]. A model based on continuous opinions takes this feature into account and may therefore implement incremental opinion changes. Importantly, if an individual shifts their opinion from 9 to 7, this shift may indeed yield implications for social influence towards others, although the qualitative opinion stance of the individual (e.g. in favor of "voting is a duty") has not changed [4].

A third type of mixed models implements aspects of both discrete and continuous frameworks. For instance, in Ref. [121], the continuous opinion of an agent i, $x_i(t)$, is computed as the average of a binary vector whose elements represent the agents' arguments in favor or against a certain issue. To mediate social influence, agents exchange such *discrete* arguments, a process which leads to *continuous* opinion changes. Another example are models of continuous opinions and discrete actions (CODA) [122, 123, 124]. Here agents decide on a binary issue or topic. However, while agents can only observe the binary actions, or opinions, of their peers with respect to the issue – as in the case of the voter model – their opinions towards the issue are not binary but continuous and lie between zero and one, i.e., $x_i(t) \in [0, 1]$. More specifically, the opinion of agent i, $x_i(t)$, is the probability that the agent chooses one side, say A, of the binary issue under consideration (A vs. B). Accordingly, for $x_i > 0.5$ agent i favors option A over B. Thus, in contrast, to the voter model, agents do not simply copy the discrete

opinions of their peers. Instead, each time an agent observes one of their peers to take action in favor of A their internal opinion $x_i(t)$ increases.

In this thesis, we will focus on agent-based models of opinion dynamics which combine two levels of description. The first level describes *who can interact with whom*. For instance, this corresponds to individuals who meet in real life or follow each other on a social media platform. The second level is concerned with the process of social influence: it specifies *how* two individuals i and j influence each other based on their opinions. To clarify why we distinguish between *interaction* and *influence* we note that the interaction between two people does generally not imply how or if they will influence each others opinions.

Taking into account all other agents in the system, one very general way to model the change of agent i's opinion, x_i, is given by

$$\dot{x}_i = \sum_{j \in \mathcal{N}_i(t)} f(x_i, x_j), \tag{1.1}$$

where we sum over the set of agents $\mathcal{N}_i(t)$ that interact with agent i at time t. The function $f(x_i, x_j)$ encodes how agent j influences agent i given there is an interaction between the two. In the model that we introduce in Ch. 7 and 8, the function f solely depends on agent j's opinion, which would correspond to $f(x_j)$ in Eq. (1.1).

In Ch. 2, we will discuss (social) networks as a way to capture the first level of description: for each agent i, the network encodes the set of their neighbors $\mathcal{N}_i$. This may either be a static set or one that evolves in time. Moreover, we will focus on models of opinion dynamics where the networks of interaction and the social influence mechanisms f are derived from, or at least motivated by, insights on human behavior. In the following section, we will therefore review some of the important previous findings.

Empirical and theoretical underpinnings of social influence

Confirmation bias

The psychological literature describes the concept of confirmation bias as the tendency of people to preferably consider information that confirms their preconceptions or reflects their existing beliefs [3, 125]. For example, people will favor a piece of information if it is consistent with their beliefs or hypotheses, otherwise they will discount, or even completely ignore it [125]. A confirmation bias can also affect people's memory. While we generally recall information which supports our existing beliefs, information supporting the conflicting side is rather easily forgotten [126]. Another striking aspect of the confirmation bias is the fact that people tend to see what they are looking for. More specifically, prior expectations strongly influence our social perception [127]. As we discuss in Ch. 6, models of opinion dynamics often implement confirmation biases as sharp thresholds in the social influence function f. In particular, these thresholds imply that the opinions of two agents i and j do not influence

each other if their opinion distance, $|x_i - x_j|$, is larger than this threshold.

Homophily

The homphily principle states that people partially form their social ties based on the similarity to their peers. There are various attributes according to which this similarity may be defined. In their seminal work on homophily P. Lazarsfeld and R. K. Merton therefore distinguished between *status* and *value homophily* [128, 129]. The former, status homophily, concerns sociodemographic dimensions such as race, gender, or age, and characteristics like education, religion, or occupation, along which society is stratified. Value homophily, on the contrary, refers to internal states, such that people with similar attitudes, beliefs and values attract each other [128]. There is extensive empirical evidence for both types of homophily. For example, homophily on race has been identified in different relationships, ranging from marriages [130] and friendships [131], to work relations [132]. Similar patterns have been observed for education, occupation, and social class. It has also been suggested that people associate with those peers that share their political orientations [133, 134], beliefs, or attitudes [135].

Note that, in this thesis we do not consider homophily as a micro-mechanism of social influence, i.e., we do not assume homophily to influence the way individuals influence each other. Instead, the assumption is that homophily determines the selection process of interaction partners. In the model introduced and discussed in Ch. 7 and 8, homophily is based on the similarity between agents' opinions. Thus, in cases of high homophily, social interactions take place more likely among people holding similar opinions on an issue.

Persuasive-arguments theory

People may change their opinions as they are persuaded by convincing arguments, an empirical observation which has been substantiated by *persuasive-arguments theory* (PAT) [30, 31]. According to PAT, a person evaluates an alternative A relatively to B, by generating a set of arguments in favor or against both options. Typically the set is finite and given by the associated cultural context. Furthermore, individual arguments are not necessarily equally important, and instead, may vary with respect to their availability and persuasiveness [31]. Finally, PAT derives the direction (A, B) and the magnitude (weak, moderate, ..., extreme) of a person's opinion from the set of their individual arguments. Intuitively, PAT suggests that social influence among individuals is a process of exchanging arguments with others, which updates one's own set of arguments [121].

Empirical evidence for persuasive-arguments theory is common, and PAT helped to explain results of different experiments on choice shifts [30, 136] and group polarization [120, 137]. Especially, group polarization will be important in Part II of this thesis, as a motivation for a simple social reinforcement mechanism. Group polarization refers to the tendency of groups to make decisions which are significantly more extreme than the initial bias, or inclination,

of the group's members, (e.g. towards greater risk or caution for initially risky or cautious inclinations, respectively). A similar phenomenon can be observed for the opinions of groups, where initial opinion biases are reinforced such that convictions are increased. The latter effect is specifically referred to as attitude polarization [138]. The opinion dynamics model, introduced in Ch. 7 and 8, is based on a simple mechanism, which mimics PAT in the sense that agents with equal stances reinforce their opinions. This may lead to a radicalization phenomenon based on the concept of group or attitude polarization.

Finally, a remark is in order. Both group polarization as well as attitude polarization should not be confused with opinion polarization. The first two concepts refer to processes, which lead to more extreme decisions or opinions in groups. In contrast, opinion polarization refers to the macro-state of a social system, where the overall opinion distribution towards a certain topic is bimodal with one peak corresponding to each opinion stance and absent, or rare, neutral opinions.

Cognitive consistency theories

The fundamental assumption in cognitive consistency theories is that people seek coherent or consistent, beliefs, attitudes, and values [35, 36]. If, however, they are experienced as contradictory, or inconsistent, this will induce an unpleasent "tension state", which individuals try to overcome.

One way to formalize cognitive consistency is by Heider's balance theory, developed in the 1940s [35]. It is based on an individual's perception on triads, formed by herself P and two further elements, O (other person or object) and X (person or object). According to balance theory, each of the three relations within a given triad, P-O, P-X, and O-X can be positive (+) or negative (−). This gives rise to eight possible combinations, where triads are unbalanced if the product of these relations is negative, and balanced if it is positive. While balanced triads are stable, unbalanced ones are experienced unpleasently due to cognitive dissonance. Thus, person P changes their opinion in order to stabilize the triad and reach a balanced state. As an example, let us look at a simple social setting involving the agents i and j and a topic k, e.g. nuclear energy. Let us furthermore assume that the agents are friends, and that agent i is in favor of nuclear energy, hence, the relations i-j and i-k are positive (+). If agent j, however, dislikes the use of nuclear energy (i.e., j-k is negative), this results in an unbalanced triad. To stabilize the triad, and to overcome cognitive dissonance, agent i has two options: (i) to quit the friendship with agent j, or (ii) to change their attitude towards nuclear energy.

2 (Social) Networks

Notably, the social sciences have the longest tradition to study networks generated from empirical data [62]. A social network describes a set of individuals and their interrelations, with respect to a wide variety of social settings, ranging from friendships [64, 139], to sexual contacts [140, 141], to business relations [142, 143]. In models of opinion dynamics, social networks often appear as the substrate of social influence and are therefore important for investigations of opinion formation on a collective level. In the following, we present an overview of relevant notations and definitions, and outline how networks can be formalized mathematically using matrices [61, 144, 145]. Furthermore, we review important characteristics of social networks, and introduce some generative network models, which will appear throughout the thesis.

Basic definitions

Formally, a network can be represented as a mathematical object called graph. A graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ defines how a set $\mathcal{N}$ of N nodes (vertices) is connected by a set $\mathcal{E}$ of N_e edges (links). Specifically, each of the $N_e = |\mathcal{E}|$ edges in the network is given as a tuple $(i, j) \in \mathcal{E}$ and encodes the relation between two nodes, a source i and target j. If such relational information is binary, the corresponding network is conveniently represented by the $N \times N$ *adjacency matrix* $\mathbf{A}$. Its elements A_{ij} are defined as

$$A_{ij} = \begin{cases} 1\,, & \text{if } (i,j) \in \mathcal{E}\,, \\ 0\,, & \text{else}\,. \end{cases}$$

By summing the adjacency matrix over the ith row or ith column, one obtains the out- or the in-degree of node i, denoted as $k_i^{\text{out}} = \sum_j A_{ij}$ and $k_i^{\text{in}} = \sum_j A_{ji}$, respectively. In directed network in- and out-degrees of nodes are generally not equal. While the in-degree (k_i^{in}) indicates from how many nodes j node i may be reached, its out-degree (k_i^{out}) counts the number of nearest neighbors in the other direction, i.e., how many nodes are reachable from node i. In this thesis, we will mainly deal with undirected networks. Here, an edge from i to j implies that there is also an edge from j to i, and hence the adjacency matrix is symmetric, with $A_{ij} = A_{ji}$, or $\mathbf{A} = \mathbf{A}^T$. In such cases, the in- and out-degree of node i are equal, and we

call $k_i \left(= \sum_j A_{ij} = \sum_j A_{ji}\right)$ simply the *degree* of node i.

The binary character of the adjacency matrix $\mathbf{A}$ prevents the encoding of more detailed information about the relation between two nodes in the network. In particular, this limits the descriptive power of the network approach, if it is important to distinguish between strong and weak couplings among nodes. Especially, with regard to social influence models it may be useful to encode e.g. the frequency or the intensity of peer interactions, which can generally not be assumed equal for all pairwise interactions [77]. Thus, a straightforward generalization of binary adjacency matrices suggests the use of a weighted coupling matrix $\mathbf{W}$. Here, the actual backbone network, encoding the binary information about whether or not an edge exists between two nodes i and j, is still given by $\mathbf{A}$, but each edge $(i, j) \in \mathscr{E}$ is characterized by a weight $W_{ij} \in \mathbb{R}$. Such weighted adjacency matrices appear in a wide range of social influence models, as we will see in Part I of this work. In the case of weighted network the definition of the node degree, also called node strength, needs to be refined. On undirected weighted networks, the strength of node i is defined as $s_i = \sum_j W_{ij}$.

Most commonly networks are described as static objects. In the context of social systems, however, this static view often yields an oversimplification, as the edges of social networks are not continuously active, or their strength varies in time [146]. For instance, person-to-person communication including e-mail messages [147, 148, 149], mobile phone text messages [150], or social exchange in online forums [151] is generally not a continuous process, but instead happens as discrete events. This property has important implications for dynamical processes sustained by these temporal interactions and motivates a time-varying representation of networks. Generally, time-varying networks come in different forms. This thesis mostly deals with two different kinds: (i) unweighted contact networks with a changing topology (or adjacency matrix $\mathbf{A}$), and (ii) networks with constant backbone $\mathbf{A}$ and time-varying weights. The former case is usually referred to as temporal network. It is defined as an ordered set $\{\mathbf{A}(t)\}$, with $t = 1, \ldots, T$, where in each time step a different adjacency matrix is realized. In the latter case the time-dependence is not due to a changing adjacency matrix. Instead, the backbone network is fixed and the coupling strengths, or weights, $W_{ij}(t)$, depend on time.

Network Laplacian

The adjacency matrix $\mathbf{A}$ (or the weighted coupling matrix $\mathbf{W}$) is not the only way to represent static networks algebraically. Another important representation is given by the network *Laplacian*, denoted as $\mathbb{L}$. As the adjacency matrix, $\mathbb{L}$ is a $N \times N$ matrix, which is defined elementwise as,

$$\mathbb{L}_{ij} = \begin{cases} -A_{ij} & i \neq j, \\ \sum_k A_{ik}, & i = j. \end{cases} \tag{2.1}$$

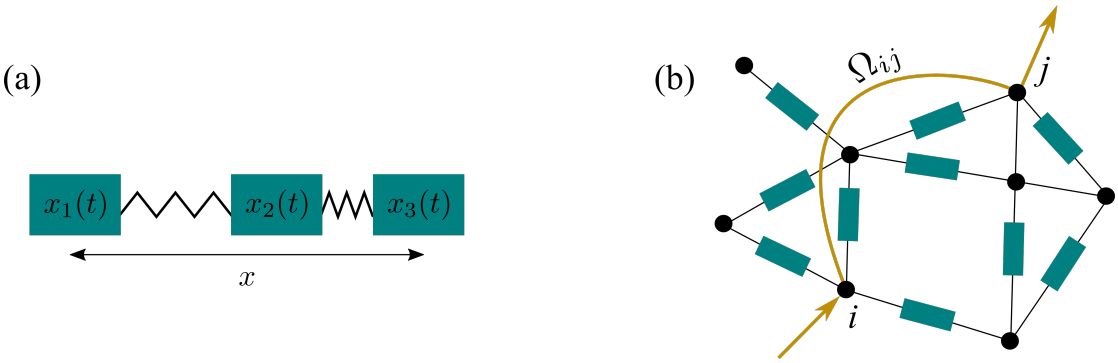

Figure 2.1: Illustration of two coupled systems involving the network Laplacian $\mathbb{L}$. Panel (a): simple mass-spring model with three coupled masses. Panel (b): electrical network, where each undirected edge is replaced by a unit resistance. The effective resistance between two nodes i and j is given as the resistance distance, as defined in Eq. (2.7).

In the case of undirected[1] networks, the Laplacian can simply be defined as $\mathbb{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal degree matrix, with $D_{ii} = \sum_k A_{ik}$. In Ch. 4 and 5, we will use a Laplacian formalism to describe coupled dynamical systems in opinion dynamics and other applications, where the underlying network is undirected and connected. In those cases, $\mathbb{L}$ is a symmetric matrix with real and non-negative eigenvalues, $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$, and eigenvectors $\mathbf{u}_\alpha$ that form an orthonormal basis, i.e., $\mathbf{u}_\alpha \cdot \mathbf{u}_\beta = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ denotes the Kronecker delta. Importantly, the degeneracy of the eigenvalue λ_1 equals the number of connected components in the network, hence, in connected networks the second smallest eigenvalue λ_2 is nonzero [152].

The network Laplacian has been utilized in many areas of network analysis [153, 154, 155], and naturally arises in various contexts of networked complex systems, ranging from random walks on networks [156, 157], to coupled oscillators for the investigation of power grid dynamics [158, 159]. In the following, we discuss two examples involving the network Laplacian, that will be instructive for the considerations of Part I.

The first example is an intuitive analogy from classical mechanics, the mass-spring model. For the ease of depiction and reasoning, we consider a very simple network, where three point masses are connected via two linear springs, as shown in Fig. 2.1(a). Combining Newton's second law with Hooke's law, and assuming unit masses and unit spring constants, the system's coupled equations equations of motion are

$$\begin{aligned} \ddot{x}_1 &= -x_1 + x_2\,, \\ \ddot{x}_2 &= x_1 - 2x_2 + x_3\,, \\ \ddot{x}_3 &= x_2 - x_3\,, \end{aligned} \tag{2.2}$$

[1] For directed networks the Laplacian is either defined in terms of k_i^{in} or k_i^{out}, depending on the application.

where x_i denotes the position of the ith mass. In vectorial form, the system of equations can be written as

$$\ddot{\mathbf{x}} = -\mathbb{L}\mathbf{x}, \tag{2.3}$$

where the network Laplacian naturally arises due to the linear coupling between the masses. Importantly, this so-called *diffusive coupling* scheme, induced by network Laplacian, is not limited to the Newtonian dynamics of networked mass-spring systems [160], or the study of other (linearized) coupled oscillator models [158, 161]. As we will discuss in Part I, various classical models of social influence, and consensus models arising in other contexts, are also based on similar coupling schemes.

The name "diffusive coupling" derives from the fact that the network Laplacian is used to model diffusion processes on networks [156, 157]. Why this is so becomes more transparent by taking a closer look at the effect of $\mathbb{L}$ on an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$. Let's assume for now that the elements of $\mathbf{x}$ do not correspond to the positions of point masses, but rather to the concentrations of a diffusing material at each node of a network $\mathbf{A}$. The ith element of the product $\mathbb{L}\mathbf{x}$ reads

$$[\mathbb{L}\mathbf{x}]_i = \sum_j \mathbb{L}_{ij} x_j = \sum_j A_{ij}(x_i - x_j), \tag{2.4}$$

which is the sum of the differences between the state of node i, x_i, and the states of all its connected neighbors x_j. According to Fick's law [162], the diffusive current between two points (here nodes i and j of a given network $\mathbf{A}$) is proportional to the corresponding difference in concentrations x_i and x_j. Hence, during a time interval $\mathrm{d}t$ the amount of material flowing from node i to one of its connected neighboring nodes j is $D(x_j - x_i)\,\mathrm{d}t$, where D denotes a rate, i.e. the diffusion constant. Summing up all contributions for node i, we get

$$\dot{x}_i = -D\sum_j A_{ij}(x_i - x_j), \tag{2.5}$$

$$= -D\sum_j \mathbb{L}_{ij} x_j, \tag{2.6}$$

which is a diffusion equation on a given network $\mathbf{A}$.

The network Laplacian has another illustrative application, which involves electrical circuits. Let us assume an undirected network, where each edge is replaced by a 1 ohm resistance, and ask: what is the total resistance between a specific pair of nodes i and j in such a network? This situation is depicted in Fig. 2.1(b). To find the effective resistance experimentally, one would attach a battery between the two nodes and measure the resulting electrical current, with the resistance R being given by Ohm's law, $R = U/I$, where U and I denote the electric potential difference and the electrical current, respectively. Klein and Randic found a connection between the Laplacian of the resistor network and the answer to the latter question [163], and thereby circumvented such cumbersome measurements. As it turns out, the "effective

resistance", or *resistance distance*, between two nodes i and j, Ω_{ij}, is given as

$$\Omega_{ij} = \mathbb{L}^{\dagger}_{ii} + \mathbb{L}^{\dagger}_{jj} - \mathbb{L}^{\dagger}_{ij} - \mathbb{L}^{\dagger}_{ji}, \tag{2.7}$$

where $\mathbb{L}^{\dagger}_{ij}$ denotes the ijth element of the Moore-Penrose pseudo inverse of the Laplacian [164]. Note that, the pseudo inverse is used because the Laplacian is singular and therefore not invertible, i.e., $\mathbb{L}^{-1}$ does not exist. Importantly, Ω_{ij} fulfills the requirements of a distance: (i) $\Omega_{ij} \geq 0$, (ii) $\Omega_{ij} = 0 \Leftrightarrow i = j$, (iii) $\Omega_{ij} = \Omega_{ji}$, and (iv) $\Omega_{ij} + \Omega_{jk} \geq \Omega_{ik}$ (triangle inequality) [163]. Conveniently, the resistance distance can also be expressed as

$$\Omega_{ij} = \sum_{\alpha=2} \frac{(u_{\alpha,i} - u_{\alpha,j})^2}{\lambda_\alpha}, \tag{2.8}$$

where $u_{\alpha,i}$ denotes the ith element of the αth eigenvector of $\mathbb{L}$ [161]. In Ch. 4, this latter formulation will be used in terms of a modified version of the resistance distance to investigate some properties of a simple model of social influence.

Some properties of (social) networks

In the year 1967, the experimental psychologist Stanley Milgram found in experiments that people could navigate letters to an unknown person within "six degrees of separation" [165], which refers to the average number of hops a letter needed to reach its final destination. Although this navigation task was achieved without a global knowledge of the underlying social network structure, Milgrim's experiments established an important feature of social networks – the *small-world property* [166, 167]. One of two characterizing features of such small-world networks is defined based on the concept of the shortest path l_{ij} between two nodes i and j, which also refers to their geodesic distance. Generally, a path Γ_{ij} from node i to node j corresponds to an ordered sets of nodes and edges, such that starting from node i, node j is reached without visiting any node twice. Hence, the shortest path between two nodes i and j is defined as the path of minimum length

$$l_{ij} = \min_{\{\Gamma_{ij}\}} \sum_{(k,l)\in\Gamma} A_{kl}. \tag{2.9}$$

Note that, for directed networks the shortest path between two nodes i and j, is not necessarily symmetrical, and we have $l_{ij} \neq l_{ji}$. In undirected networks, averaging the shortest paths l_{ij} (Eq. 2.9), between all possible pairs of nodes i and j, yields

$$\langle l \rangle = \frac{2}{N(N-1)} \sum_{i<j} l_{ij}, \tag{2.10}$$

which quantifies the expected geodesic distance (or shortest path) between two nodes in a given network $\mathbf{A}$. Crucially, in small-world networks, $\langle l \rangle$ is found to be small compared to size of the network, and it scales logarithmically (or slower) with the number of nodes, i.e.,

$\langle l \rangle \sim \ln(N)$ [66].

In addition to small shortest path lengths, small-world networks are characterized by large clustering coefficients [144]. The local clustering coefficient c_i of node i is defined as

$$c_i = \frac{2e_i}{k_i(k_i-1)}, \tag{2.11}$$

where e_i denotes the number of edges between the neighbors of node i. In particular, Eq. (2.11) defines c_i as the ratio between the number of actual connections between agent i's neighbors (e_i), and the maximum possible number of such connections, given as $k_i(k_i-1)/2$. Averaged over all nodes of the network, c_i gives rise to the global clustering coefficient $\langle c \rangle = N^{-1}\sum_i c_i$. The concept of clustering is very similar to the definition of transitivity in sociology, which quantifies the tendency of a social network to form local cliques [144, 168]. Accordingly, if **A** encodes a friendship network, high clustering coefficients suggest that most of a person's friends will also be friends.

Clustering in networks is connected to a further structural feature present in many real-world social networks, namely community structure [169, 170]. Although the term network community is somewhat loosely defined, it generally describes subsets of network nodes, connected more densely than suggested by a random arrangement of edges [170]. In particular, if the number of edges within certain groups of nodes is high, while the number of edges between those groups is rather low, those node groups may correspond to communities in a given network. To quantify this concept one may define the modularity of a network, Q, as

$$Q = \frac{1}{4N_E}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2N_E}\right)\delta_{\mathrm{com}_i,\mathrm{com}_j}, \tag{2.12}$$

where com_i is the community identifier of node i, such that $\delta_{\mathrm{com}_i,\mathrm{com}_j} = 1$ if nodes i and j are in the same community, and $\delta_{\mathrm{com}_i,\mathrm{com}_j} = 0$ otherwise. Hence, the expression for Q compares the actual number of edges among two nodes i and j in a given community A_{ij} (which is zero or one for unweighted networks), with the number of edges expected in the case of a random graph $k_i k_j/(2N_E)$. Identifying the community structure of a network means to partition nodes into groups (the communities), which maximize the value of Q, as defined in Eq. (2.12). Different algorithms, based modularity maximization, have been proposed and applied to the analysis of various social networks [170, 171, 172, 173], one of which will be used in Ch. 8 to identify network communities in a model of opinion dynamics.

From a statistical perspective, the most basic and straightforward way to characterize networks is by their degree distribution $p(k)$. For undirected networks[2], the degree distribution is defined as the probability $p(k)$ that a randomly chosen node is of degree k. Empirically, most networks can be divided into two broad classes; those with homogeneous, and those with heterogeneous degree distribution [144]. In the first case, degrees are distributed according to

[2] In the case of directed networks it is distinguished between in-degree and out-degree distributions.

a Poisson or a Gaussian distribution, which have a light, exponential tail. Interestingly, many real-world networks do not fall into this first class of homogeneous degree network. Instead, their degree distribution is a heterogeneous one, usually following a power law $p(k) = Ak^{-\gamma}$, with exponent γ. Due to their power law degree distributions, those networks are referred to as *scale-free* [73]. The implications of power law degree distributions on the dynamical processes running on top of scale-free networks has been subject to extensive research. Importantly, it was found that due very strongly connected nodes, so-called hubs, the spreading of epidemics, computer viruses, and cultural items behave fundamentally different as in homogeneous random networks [76, 174]. In the next section, we will discuss different network generating models, which give rise to both types of degree distributions.

(Social) network models

Besides the analysis of empirical (social) networks, a second important branch of network science is concerned with the development of models, which are based on simple principles that generate networks with similar properties to the ones found in empirical studies. In the following, we will briefly review some important works, with a focus on those network models, which will be relevant for the following chapters of this thesis.

Erdős-Rényi model

The simplest model to describe random networks of arbitrary size was developed by P. Erdős and A. Rényi in 1959 [144, 175]. The idea behind Erdős-Rényi (ER) networks is simple. Let us consider a set of N nodes, where each of the possible $(N-1)N/2$ (undirected) edges is independently realized with probability p_{ER}. The result is a random network, with an expected number of $\langle N_E \rangle = p_{\mathrm{ER}} N(N-1)/2$ edges. From these assumptions, it follows that, the degree distribution of ER networks is the binomial distribution

$$p(k) = \binom{N-1}{k} p_{\mathrm{ER}}^k (1-p_{\mathrm{ER}})^{N-1-k}, \tag{2.13}$$

which yields an average degree and average clustering coefficient of $\langle k \rangle = (N-1)p_{\mathrm{ER}}$ and $c(k) = \langle k \rangle / N$, respectively [144]. For $N \gg \langle k \rangle$ the binomial degree distribution [Eq. (2.13)] is well described by a Poissonian one, i.e. we have $p(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$. Thus, while ER networks fulfill the first criterion of small-world networks, namely the logarithmic scaling of average shortest paths $\langle l \rangle$ with the system size, they show high clustering coefficients $\langle c \rangle$, only in the limit of $p_{\mathrm{ER}} \to 1$.

In Fig. 2.3(b), we show an ER network for $N = 100$ and $p_{\mathrm{ER}} = 0.05$. Below we depict the resulting degree distribution of a larger network of $N = 1000$ nodes. The overall distribution is very well captured by the Poissonian distribution, which is depicted as red solid line.

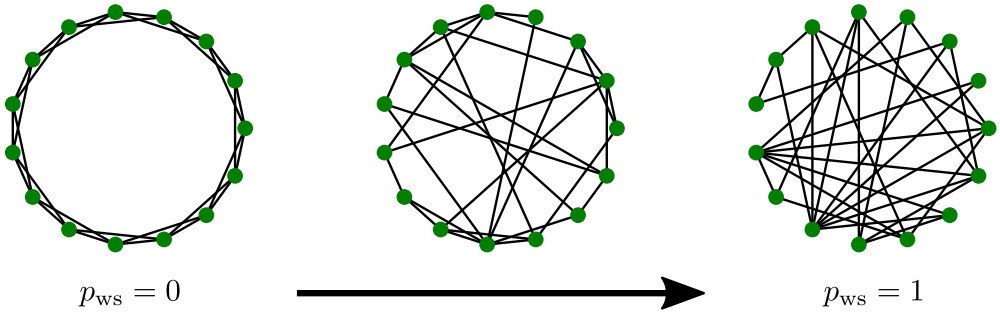

Figure 2.2: Schematic depiction of the WS model for a network with $N = 15$ nodes and $k_{\mathrm{ws}} = 4$ for different values of p_{ws}: 0 (right), 0.5 (center), and 1 (right).

Watts-Strogatz model

For long it was suggested that both criteria of small-world networks, i.e., short average path lengths $\langle l \rangle$ and high clustering $\langle c \rangle$, are mutually exclusive. On the one hand, there are networks such as ring lattices, which have high clustering coefficients in combination with very large average shortest path lengths [66]. On the other hand, there are random networks, as described by the ER model, which yield the desired logarithmic scaling of $\langle l \rangle$ with the system size, however, do not posses high clustering coefficients, which are characteristic for many real-world networks [144]. This view was rejected by the seminal work of Watts and Strogatz [66], where they emphasized that many real networks, such as e.g. collaboration networks, combine both features.

Accordingly, their idea was to design a model which may interpolate between these two structures. The algorithm to generate a Watt-Strogatz (WS) network starts with a ring of N nodes, where each node is connected to its k_{ws} nearest neighbors. Subsequently, shortcuts are added to the network as follows: each edge (i, j) is replaced by a new edge (i, k) with probability p_{ws}, where the new neighbor of node i, node k, is chosen uniformly random over the entire ring, such that duplicate edges are forbidden. In Fig. 2.2 we depict three WS networks with $N = 15$ nodes and increasing values of $p_{\mathrm{ws}} = [0, 0.5, 1]$ in ascending order from left to right, interpolating from a perfectly ordered grid to a random network with similar statistical properties as an ER network.

The surprising insight was that the average shortest path $\langle l \rangle$ of networks is drastically reduced already for very small values of p_{ws}. At the same time, the clustering coefficient remains high due the lattice backbone of the network, such that both features of small-worldness can be observed simultaneously over a wide range of p_{ws} values.

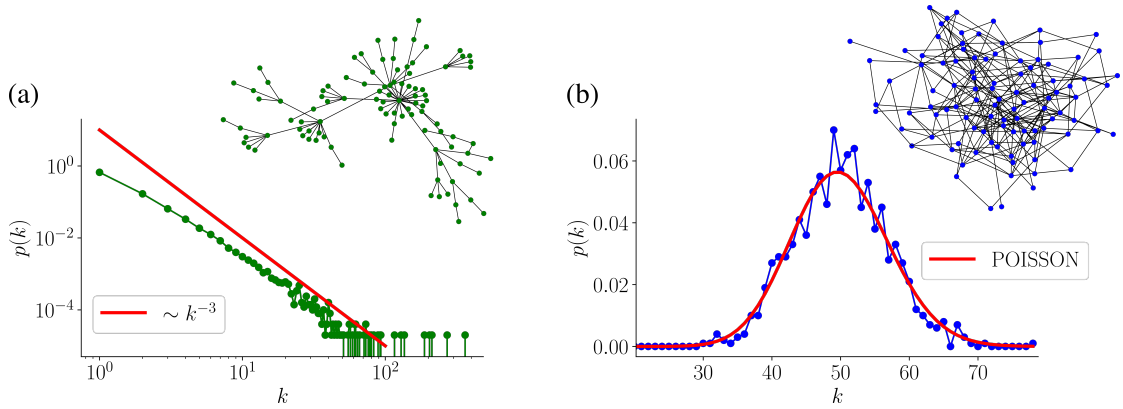

Figure 2.3: Examples of a BA network [panel (a)] and a ER network [panel (b)] and typical degree distributions of the corresponding network models. Both depicted networks contain $N = 100$ nodes, we have set $m = 1$ and $p = 0.05$ for the BA network and the ER network, respectively. While the degree distributions correspond to much larger networks ($N = 50000$ for BA and $N = 1000$ for ER) the values of m and p have not been changed.

Barabási-Albert network

Despite their great success in describing important features of various real-world networks, both the ER and the WS model do not capture power law degree distributions, which have been observed in various systems, ranging from biological [176, 177], to technological [73], to many social networks [68]. Importantly, a power law degree distribution implies that, while the majority of nodes has a small degree, a small number of nodes has very high degrees. Interestingly, power laws statistically characterize a wide range of other phenomena [178] including the sizes of cities [179], the frequencies of words used in human languages [180], the market shares of brands [181], and people's annual incomes [182] suggesting similar mechanism underlying these processes.

The first model to explain the emergence of scaling in networks' degree distributions was proposed by L. Barabási and R. Albert in 1999 [73]. The model is based on two realistic mechanisms: network growth and preferential attachment [183]. First, the algorithms starts with a small seed network, which is usually assumed to be fully connected. Subsequently, at each time step t a new node enters the network, until the final network size N is reached (growth). Crucially, each new node chooses, upon entry, m_{BA} other nodes j to which it connects with probability $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$, i.e., new nodes preferentially pick nodes with high degree (preferential attachment).

This process results in a linear growth of the number of nodes and edges as $N(t) \sim t$ and $N_E(t) \sim m_{\mathrm{BA}} t$, respectively, and yields the power law degree distribution

$$p(k) = \frac{2m_{\mathrm{BA}}(m_{\mathrm{BA}}+1)}{k(k+1)(k+2)}, \tag{2.14}$$

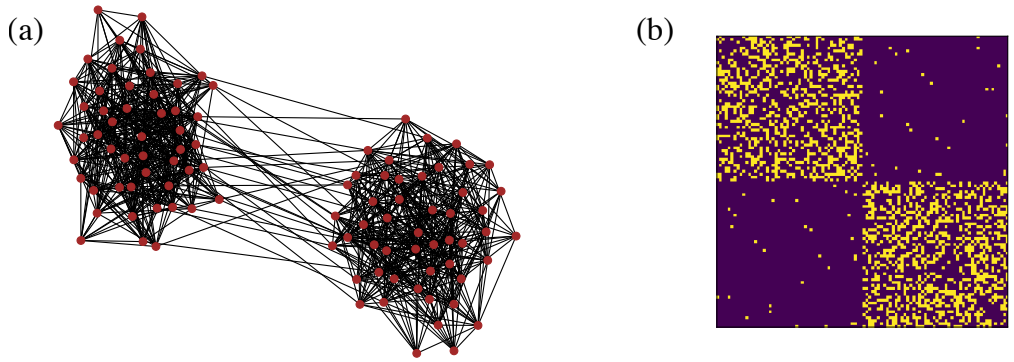

Figure 2.4: A random network generated by the stochastic block model and the corresponding adjacency matrix depicted as a heatmap are shown in panels (a) and (b), respectively. The network consists of $N = 100$ nodes and two blocks ($b = 2$). While the probability for a link between two nodes in the same block is set to a rather large value of $p_{\text{intra}} = 0.35$, the inter-group connection probability is low ($p_{\text{inter}} = 0.01$). This choice of parameters results in two pronounced communities, as clearly apparent in the network depiction [panel (a)]. In a different representation the community structure is clearly visible in a heatmap of the symmetric adjacency matrix, where each yellow cell correspond to a directed link in the network.

which scales as $p(k) \sim k^{-3}$ for high degrees [184].

In Fig. 2.3(a), we show a BA network for $N = 100$ and $m_{\text{BA}} = 1$. Below, the degree distribution of a much larger BA network of $N = 50000$ (and $m_{\text{BA}} = 1$) is depicted. In contrast to the (Poissonian) degree distribution of an ER network [see panel (b)], $p(k)$ of the BA network does not obey a typical scale and follows a power law, as predicted by Eq. (2.14), with $p(k) \sim k^{-3}$ for $k \gg 1$ (red solid line).

Stochastic block model

A straightforward way to obtain random networks characterized by a pronounced community structure is provided by stochastic block models (SBM) [185]. Here, the nodes of a network are separated into discrete groups, so-called blocks. In its standard formulation the SBM assumes that all nodes which belong to one block are statistically equivalent, and that a connection between two nodes is solely determined by their group membership [186, 187].

More specifically, the model is defined as follows. Each of the N nodes in the network is assigned to one of b blocks, and n_r is the number of nodes in block $r \in [0, b-1]$. Then a $b \times b$ matrix, P_{SBM}, is defined, whose element $[P_{\text{SBM}}]_{kl}$ corresponds to the probability for an edge to exist between any node i in block k, and any node j in block l. Note that if $[P_{\text{SBM}}]_{kl} = p \; \forall \; k, l$, the SBM reduces to the ER network.

In Fig. 2.4, we show a SBM network with $N = 100$ nodes [panel (a)], and the heatmap of the

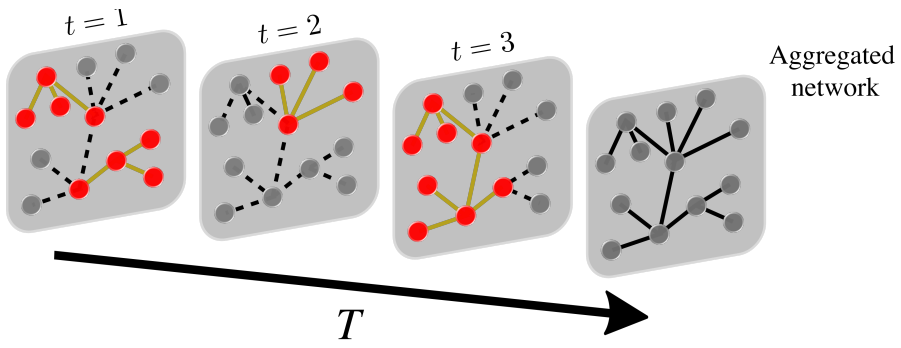

Figure 2.5: Schematic representation of the AD model with $N = 13$ and $m = 3$ for three times steps. The active nodes in each time step and the nodes which they connect to are colored in red. In-active nodes are shown in grey. Links generated by active nodes are shown in dark yellow. The aggregated network contains all links generated during the time course (here consisting of three network snapshots) of the AD network dynamics.

corresponding adjacency matrix, where cell (i, j) corresponds to the matrix element A_{ij} and is colored in yellow if there is an edge from node i to j [panel (b)]. For simplicity, we consider a network of two blocks with $n_0 = n_1 = 50$, where the probability for links between nodes within one block ($p_{\text{intra}} = [P_{\text{SBM}}]_{11} = [P_{\text{SBM}}]_{22}$) is much larger, than the probability for links between nodes in different blocks ($p_{\text{inter}} = [P_{\text{SBM}}]_{12} = [P_{\text{SBM}}]_{21}$). This choice results in a pronounced community structure, clearly visible in both the network depiction as well as in the heatmap of the adjacency matrix. A similar setup will be used in Ch. 4 in order to investigate the effects of community structure on the resulting opinion dynamics within a simple model of social influence.

Activity-driven network model

The activity-driven (AD) framework provides a simple framework to model Markovian temporal networks [59]. It has been used to investigate the impact of network temporality on a variety of different dynamical processes, ranging from random walks [188, 189, 190], to epidemic spreading [191, 192, 193, 194], to opinion dynamics [105, 106, 195, 196].

In contrast to all previously discussed models, whose definition is based on the (static) connectivity of a network, the AD model focuses instead on the activity of individual agents. Those agents are assumed to interact in discrete time steps, forming temporal connections. Thus, the network is not given by a static but a time-varying adjacency matrix $\mathbf{A}(t)$.

Formally, the AD model is defined as follows. Each of the N agents in the system is assigned a static activity value $a_i \in [\varepsilon, 1]$, which is drawn from a given distribution $F(a)$. Typically, $F(a)$ is assumed to follow a power law as $F(a) \sim a^{-\gamma}$ with exponent γ and a lower activity bound $\varepsilon > 0$ to avoid divergence. Then, in each discrete time step t, the instantaneous network $\mathbf{A}$ is initialized as a set of disconnected nodes. An agent is activated with a probability proportional

to their activity a_i. Each active agent generates m links to randomly sampled other nodes in the network, where non-active nodes can still receive links. In the following time step $(t+1)$, all previous links are deleted and the activation-linking process starts anew, thus rendering the process Markovian. In Fig. 2.5, we depict a schematic representation of the AD model with $N = 13$ and $m = 3$ for three network snapshots. In each time step, active nodes, and the nodes they connect to, and the corresponding links are colored in red and yellow, respectively. Inactive nodes are shown in grey. The aggregated network, shown to the right, contains the links generated in all previous time steps.

In its original definition, and with respect to most of its previous applications [105, 106, 188, 189, 190, 191, 192, 193, 194, 195, 196], the AD model differs from adaptive network models, where the network dynamics is coupled to a dynamical process running on top of the network [197]. By contrast, in Ch. 7 and 8, we integrate the AD model into a modeling framework for opinion dynamics, where the temporal network and the opinion dynamics are coupled via a simple homophily mechanism and therefore co-evolve.

Classical Consensus Models Part I

3 Consensus by Assimilative Influence

Social influence has been found to often reduce the differences between individuals, and various *assimilative* processes have been identified in different social contexts [3, 4]. For instance, the experiments performed by S. E. Asch in 1956 showed that individuals tend to conform with a majority group [198]. Other examples include research on persuasion [4, 30, 31], and social learning [32], which can also explain why people may become more similar upon social exchange [32]. While in persuasion research one assumes that people may assimilate due to the exchange of arguments, social learning theories emphasize other factors, such as imitating behavior [32]. Other possible assimilation mechanisms can be derived from cognitive consistency theories [35], where disagreement may lead to an unpleasant cognitive dissonance, which people aim to overcome by adapting to their social environment.

In Part I of this thesis, the focus is on formal models of assimilative influence, which give rise to states of consensus. Below, we will introduce such consensus models from an opinion dynamics perspective. Subsequently, we discuss their use also outside of social dynamics in more technical applications. Broadly, opinion assimilation defines a process in which agents' opinions *always* become more similar as a result of social influence [3]. While their implementation differs across the literature, here, we focus on assimilative models where agents hold continuous opinions and their pairwise interactions are encoded in a social influence network. Traditionally, opinion assimilation was implemented in discrete time with agents updating their opinions in an iterated averaging procedure involving the opinions of their connected peers. By contrast, we will focus on consensus models in continuous time, which are based on a diffusive coupling scheme induced by the Laplacian of the underlying social influence network. In our contributions, discussed in Ch. 4 and 5, we investigate two different consensus models. While the first one is a model for opinion dynamics, the second one has been applied in more technical domains, as in studies of multi-vehicle formation. Both models are investigated within a spectral decomposition approach in order to analyze analytically different mechanisms inhibiting the formation of collective consensus states.

3.1 Modeling (social) assimilation

The earliest formalization of assimilative social influence on networks dates back to J. R. French's work "*A formal theory of social power*" [81]. It can be seen as the founding manifesto of social influence network theory – the study of how social influence and the structure of the influence network shape the collective opinion dynamics on the group level. Later, the idea of formalizing social influence mathematically was developed further by scientists from with diverse background, including, amongst others, mathematicians, social psychologists, and physicists [4, 38, 77].

French assumed that individuals are embedded in a network of social (interpersonal) influences. This network approach can be motivated by the idea that individuals are not only influenced by one but potentially various distinct peers. By repeatedly shifting their opinions toward the mean opinion of their peers, French hypothesized, individuals balance tensions resulting from disagreements in their interpersonal relationships [77]. While the opinion of each agent is assumed to vary on a continuous scale, the model is implemented in discrete time, leading to an iterated averaging procedure. In particular, the time evolution of the opinions of N agents is given by

$$\mathbf{x}(t+1) = \mathbf{W}\mathbf{x}(t)\,, \tag{3.1}$$

where the ith element of vector $\mathbf{x}(t) \in \mathbb{R}^N$, $x_i(t)$, represents the opinion of agent i at time t and time, $t = 1,2,3\ldots$, is counted in discrete steps. The matrix $\mathbf{W}$ encodes the weights of interpersonal influences. In particular, the weight W_{ij} denotes the influence of agent j on agent i, defined as

$$W_{ij} = \frac{A_{ij}}{\sum_k A_{ik}},$$

where the time-independent matrix $\mathbf{A}$, with elements A_{ij}, is the adjacency matrix of the underlying, static social influence network. Thus, $A_{ij} = 1$ means that agent j's opinion influences agent i's opinion, otherwise it is assumed that $A_{ij} = 0$. The influence is reciprocal, i.e., agent i influences agent j, if $A_{ji} = 1$. Note that, in the French model it always holds that $A_{ii} = 1\ \forall i$, which results in non-vanishing self-weights $W_{ii} > 0$ for all agents. Agents' self-weights, or resistances, have two intended consequences. First, agents do not only take into account the opinions of their neighbors in the influence network, but also their own opinions. Second, the opinions of agents on which no social influence is exerted remain constant; without the self-weight $W_{ii} = 0$ the opinion of a disconnect agent i would vanish after a single iteration of Eq. (3.1). Broadly, the French model formalizes mathematically, in terms of a simple iterative averaging mechanism, assimilative social influence, where agents cognitively integrate different influences of their connected peers, based on a given social influence network $\mathbf{W}$ [77]. Importantly, this averaging procedure, which implies that the opinions of individuals move toward each other, has been empirically verified for discussants in experiments of dyadic interactions [199, 200]. In Fig. 3.1, the resulting opinion dynamics

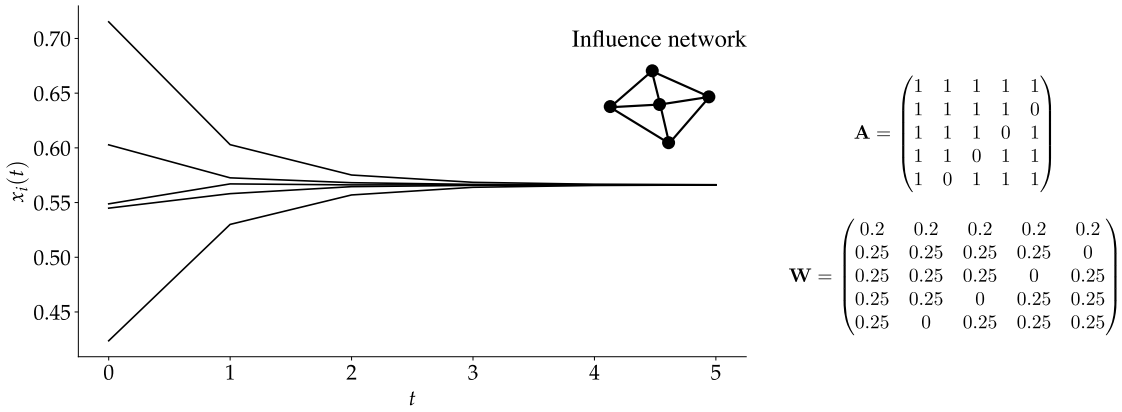

Figure 3.1: Opinion evolution in discrete time steps according to French's model of iterated opinion averaging as defined in Eq. (3.1) and (3.2). While the backbone network is undirected, i.e., **A** is symmetric, the matrix encoding the influence weights **W** is not symmetric, due to the row-normalization condition of the French model, see Eq. (3.2).

according to the French model is illustrated. While the backbone network is assumed to be undirected (**A** is symmetric), the matrix encoding the weights of social influence, **W**, used for the iterative opinion averaging procedure, is not symmetric due to the row-normalization condition.

Several generalizations of French's opinion averaging model were introduced. Specifically, we mention the models of Harary [82] and DeGroot [83], due to their strong similarity to French's approach [77]. While in those models the collective opinion dynamics follows Eq. (3.1), the definition of the influence matrix **W** differs from French's model. First, Harary relaxed the assumption that a person's self-weight, W_{ii}, equals the influences of their connected peers [82]. Independently of Harary, DeGroot went a step further and relaxed also the homogeneity assumption, i.e., that agents are generally influenced equally by all their connected peers in the social influence network [83]. Thus, DeGroot's condition on **W** simply requires that

$$0 \leq W_{ij} \leq 1 \quad \text{and} \quad \sum_{k=1} W_{ik} = 1 \;\forall\; i, j\,; \tag{3.2}$$

the row normalization of the weighted influence network **W**.

3.2 Iterated averaging and Laplacian opinion coupling

The models of French [81], Harary [82], and DeGroot [83] are formulated as iterative averaging procedures. Accordingly, the updated opinion of agent i at time $t+1$ is given as

$$x_i(t+1) = \sum_j W_{ij} x_j(t), \quad \text{for} \quad i = 1, \ldots, N. \tag{3.3}$$

This discrete-time formulation appears as a natural one, especially in the context of social psychological experiments, where the opinions of individuals may only be assessed at certain time points. From a modeling perspective, however, it is desirable to work with a continuous-time analogue of Eq. (3.3), which is formulated as a differential equation. This may especially be justified if the time elapsed between two steps of the opinion iteration is very small [200].

Subtracting the opinion of agent i at time t, $x_i(t)$ from both sides of Eq. (3.3), and using DeGroot's row-normalization condition of $\mathbf{W}$ $\left(\sum_j W_{ij} = 1 \forall i, j\right)$, yields

$$\begin{aligned} x_i(t+1) - x_i(t) &= -\sum_j W_{ij} x_i(t) + \sum_j W_{ij} x_j(t) \\ &= -\sum_j W_{ij} \left(x_i(t) - x_j(t)\right). \end{aligned} \tag{3.4}$$

To obtain a differential equation for $x_i(t)$ from Eq. (3.4), we assume that W_{ij} is not merely an influence weight. Rather it is interpreted as a (contact) rate in the spirit of a master equation [201], where the "opinion flux" from agent j to agent i, within a time-interval $\mathrm{d}t$, is given as $W_{ij}\mathrm{d}t$. Substituting $W_{ij} \to W_{ij}\mathrm{d}t$, and dividing Eq. (3.4) by $\mathrm{d}t$, yields

$$\frac{x_i(t+\mathrm{d}t) - x_i(t)}{\mathrm{d}t} = -\sum_j W_{ij} \left(x_i(t) - x_j(t)\right), \tag{3.5}$$

which offers an alternative interpretation of the averaging process defined in Eq. (3.3). The change of an agent i's opinion, during the time interval $[t, t+\mathrm{d}t]$ is given by the sum of changes towards all connected neighbors j [200], where both the magnitude and the direction of change at time t depend on the pairwise differences $(x_i(t) - x_j(t))$. In the limit of $\mathrm{d}t \to 0$, Eq. (3.5) yields the differential equation

$$\dot{x}_i(t) = -\sum_j W_{ij} \left(x_i(t) - x_j(t)\right), \quad \text{for} \quad i = 1, \ldots, N, \tag{3.6}$$

governing the time evolution of $x_i(t)$. In vectorial form, Eq. (3.6) can be expressed as

$$\dot{\mathbf{x}}(t) = -\mathbb{L}\mathbf{x}(t), \tag{3.7}$$

where $\mathbb{L}$ is the Laplacian of the weighted network, i.e.,

$$\mathbb{L}_{ij} = \begin{cases} -W_{ij}, & i \neq j, \\ \sum_k W_{ik}, & i = j, \end{cases} \tag{3.8}$$

in line with the definition of the network Laplacian introduced in Eq. (2.1).

The continuous-time perspective on opinion assimilation models was initially provided by R. Abelson, and thus Eq. (3.6) is also referred to as the *linear* Abelson model [202]. As in Eq. (3.1), it is assumed that the influence matrix $\mathbf{W}$ is non-negative to exclude repulsive interactions

between agents' opinions[1]. However, in contrast to (discrete-time) opinion averaging models the influence matrix of the Ableson model is not necessarily (row-)stochastic, i.e., generally we have $\sum_j W_{ij} \neq 1$ [84, 200]. In the following, as in Ch. 4 and 5, we assume that the (social) influence matrices, $\mathbf{W}$, are symmetric and the underlying backbone networks are connected. In this case, Eq. (3.7) always leads to a global consensus in the limit of for $t \to \infty$, which is defined as an opinion state with $|x_i(t \to \infty) - x_j(t \to \infty)| = 0 \; \forall i, j$ [203].

The emergence of consensus becomes particularly evident within a spectral Laplacian approach. The time-dependent solution of Eq. (3.7) can be decomposed as

$$\mathbf{x}(t) = \sum_\alpha c_\alpha(t)\, \mathbf{u}_\alpha \,, \tag{3.9}$$

involving the eigenvectors of the Laplacian, $\mathbf{u}_\alpha$, where $c_\alpha(t)$ denotes the αth time-dependent coefficient of this expansion. Plugging the ansatz (3.9) into Eq. (3.7) yields

$$\sum_\alpha \dot{c}_\alpha \, \mathbf{u}_\alpha = -\sum_\alpha c_\alpha \, \mathbb{L}\mathbf{u}_\alpha = -\sum_\alpha c_\alpha \, \lambda_\alpha \mathbf{u}_\alpha, \tag{3.10}$$

where λ_α is the αth eigenvalue of $\mathbb{L}$. Multiplying Eq. (3.10) with an arbitrary eigenvector $\mathbf{u}_\beta$ and performing the summations yields

$$\dot{c}_\alpha = -\lambda_\alpha \, c_\alpha \tag{3.11}$$

for each mode $\alpha = 1, \ldots, N$, where we have used that the eigenvectors of $\mathbb{L}$ are orthonormal. The general solution of Eq. (3.11) reads

$$c_\alpha(t) = c_\alpha(0) e^{-\lambda_\alpha t}. \tag{3.12}$$

In undirected and connected networks it additionally holds that, $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1}$. Therefore, in the limit of $t \to \infty$ only the zeroth eigenvalue gives a finite contribution to the stationary opinion state and we have

$$c_\alpha(t \to \infty) = c_\alpha(0) \delta_{0,\alpha} \,. \tag{3.13}$$

From the latter equation it follows that the stationary opinion state is

$$\mathbf{x}(t \to \infty) = \sum_\alpha c_\alpha(t \to \infty) \delta_{0,\alpha} \mathbf{u}_\alpha = c_0(0) \mathbf{u}_0 \,, \tag{3.14}$$

with $c_0(0) = \sum_i x_i(0)/\sqrt{N}$, where it was used that the eigenvector associated to the zeroth eigenvalue is $\mathbf{u}_0 = (1, 1, \ldots, 1, 1)/\sqrt{N}$. Therefore, the final opinion state is given as $x_i^\infty = N^{-1} \sum_i x_i(0), \forall i$, i.e., in the steady state all agents hold the average initial opinion. From this general property of the network Laplacian's spectrum follows that, given an undirected and connected network, the final stationary state is *not* influenced by the structure of the net-

[1] In Ch.6 we will discuss (non-linear) extensions to classical opinion assimilation models, where negative influence weights ($W_{ij} < 0$) are implemented to yield differentiation mechanisms between agents.

work. However, as we will see in Ch. 4 and 5 the network structure indeed critically determines important properties of non-consensus states in assimilative consensus models.

The inevitable emergence of a global consensus state for assimilation models on undirected and connected influence networks[2] leads us back to a (paraphrased) version of Abelson's question: how can one explain opinion heterogeneity [84]? Besides disconnected networks, another potential cause for the emergence and the persistence of heterogeneous non-consensus states is given by external forces acting on a subset of agents in the system. A model to investigate such effects, was proposed by M. Taylor [202]. The Taylor model extends the Abelson model, defined in Eq. (3.6), by an additional term modeling a subset of agents, which are influenced by external communication sources $s_1, \ldots, s_M \in \mathbb{R}$. The opinion evolution of agent i is governed by the following set of equations

$$\dot{x}_i(t) = -\sum_j W_{ij}\left(x_i(t) - x_j(t)\right) + \sum_k^m C_{ik}(s_k - x_i(t)), \quad \text{for} \quad i = 1, \ldots, N, \tag{3.15}$$

where $\mathbf{C}$ is a non-square and non-negative matrix $\mathbf{C} \in \mathbb{R}^{N\times M}$. The elements of $\mathbf{C}$, $C_{ij} > 0$, encode the influence of a communication source j on the opinion of agent i. Note that, an agent i with $\sum_k C_{ik} = 0$ is not influenced by any external communication source. In Ch. 4, we will investigate the emergence of non-consensus states within a formally less general version of the Taylor model [202], where each agent is at most influenced by one communication source only. Such agents will be referred to as stubborn agents, since the communication sources bias them towards certain opinions.

3.3 Consensus formation in other contexts

The formation of consensus based on assimilation is not only a well-studied phenomenon in models of social dynamics, where consensus refers to a state, in which individuals agree on a certain topic [3], share an equal set of cultural norms [97], or use the same set of vocabularies in order to communicate [204]. Consensus formation is also key for the proper functioning of many other systems, ranging from biological to artificially designed systems [204, 205, 206, 207, 208, 209, 210].

Indeed, simple assimilation mechanisms have been applied in various models of collective dynamics, in which units, coupled by an interaction network, exchange information to form a consensus in a decentralized manner. Information flows may correspond to animals exchanging cues on prey or predators [205, 206] or communicating robots within vehicular platoons [207, 208, 209, 210]. Generally, the outstanding role of consensus formation for a wide range of systems is due to collective decision making, which often requires overall agreement to achieve a collective goal [20]. For example, fish schools have shown to achieve increased levels of sensitivity if individual fish are more strongly aligned [211]. States of consensus are also

[2] On directed networks, consensus may not emerge if the network does not have a directed spanning tree [203].

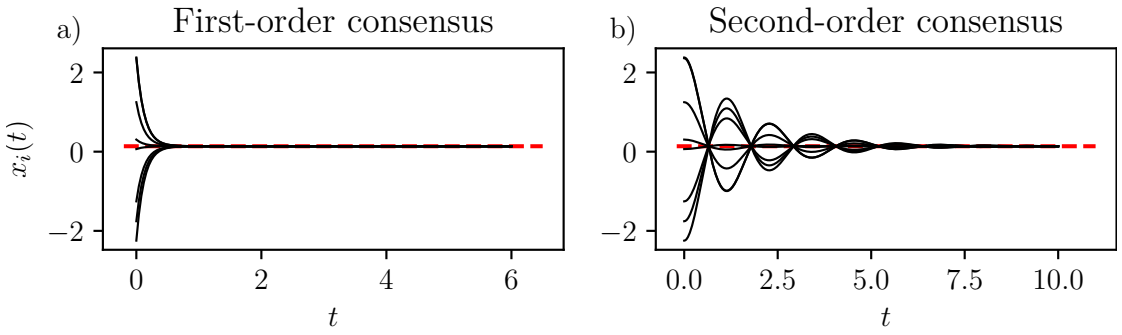

Figure 3.2: Typical transient dynamics towards a global consensus state in first- and second-order consensus models. Panels a) and b) show the trajectories of all agents (black lines) for first- (Eq. (3.7)) and second-order consensus formation (Eq. (3.16), with $\mu = 0.15$), respectively. For ease of comparison the systems are initialized with equal state variables x_i, and the initial velocities are set to zero for the second-order model, i.e. $\dot{x}_i = 0,\ \forall i$.

desirable in many technical applications. For instance, in vehicular platoons, such states may refer to situations in which all involved robots are located at the same position, point in the same direction, or move at the same speed [207, 208, 209, 212, 213]. To model consensus formation in such systems, frameworks equivalent to the linear Abelson model have been considered. They are typically referred to as single-integrator dynamics or first-order systems (see [212, 214] for extensive reviews), where the names derive from the first-order time derivative in Eq. (3.7).

While the use of first-order models may be well justified in many cases, it is often not sufficient if the inertia of individual units, agents, or robots cannot be neglected. To describe more realistically the formation of consensus among physical objects second-order models, also called double integrator dynamics, are commonly used. Here, the Newtonian dynamics of single units is taken into account by a second-order time derivative [212, 215, 216, 216, 217, 218, 219, 220, 221, 222, 223, 224], cf. Eq. (2.3) in Ch. 2. On top of the Laplacian coupling between the positions of units (x_i) a velocity alignment mechanism is implemented through the diffusive coupling between the units' velocities ($\dot{x}_i$). A general second-order consensus model on networks, including dissipation, can be formulated as

$$\ddot{x}_i + d\dot{x}_i = -\gamma \sum_j W_{ij}(x_i - x_j) - \mu \sum_j W_{ij}(\dot{x}_i - \dot{x}_j), \quad \text{for} \quad i = 1, \ldots, N, \tag{3.16}$$

where $\gamma > 0$ and $\mu > 0$ denote the coupling strengths between the agents' positions and velocities, respectively, and $d > 0$ is the damping coefficient. In contrast to first-order systems, consensus in second-order systems is defined in terms of positions (x_i) *and* velocities ($\dot{x}_i$), with $|x_i - x_j| = 0$ and $|\dot{x}_i - \dot{x}_j| = 0\ \forall\ i, j$. Note that, second-order consensus does generally not imply that velocities $\dot{x}_i$ vanish for $t \to \infty$ [212].

Figure 3.2 illustrates the dynamics towards consensus in first- and second-order systems on

the same undirected influence network. While agents approach consensus monotonically in first-order models [panel (a)], the transient dynamics of the second-order model may yield over- and undershoots, due to agents' finite inertia. In Ch. 5, we study the dynamics of Eq. (3.16) on different network topologies, where we will assume additionally that the coupling strengths are time-dependent ($W_{ij}(t)$) and periodically modulated. This type of time-varying coupling gives rise to a mechanism, which inhibits the formation of second-order consensus.

4 A Laplacian Approach to Stubborn Agents in Social Networks

This chapter is mainly based on the publication

Fabian Baumann, Igor M. Sokolov and Melvyn Tyloo.
A Laplacian approach to stubborn agents and their role in opinion formation on influence networks.
Physica A: Statistical Mechanics and its Applications 557 (2020): 124869.

We thank Elsevier for the kind permission to reuse the contained figures for this dissertation.

Established social conventions are often not persistent in real social systems. Instead, it has been suggested that they are sometimes subject to rapid changes, which may be induced by a small number of committed individuals [225, 226], a hypothesis which is backed up by recent empirical findings. In particular, in Ref. [227] the problem of social tipping points was investigated experimentally within a canonical coordination task, where people were asked to give a name to a picture of a fictitious person. Once a social convention was established, a rather small committed minority succeeded to overcome the equilibrium state, which has been established across the group.

In a similar manner, it is reasonable to assume that processes of opinion formation may be strongly altered by subsets of individuals who are committed to a certain opinion, or develop incentives to deviate from an established consensus. In particular, recent developments on social media platforms may prompt researchers to address such issues. This may include bots (short for software robot) populating such social media platforms, where they have been used to hijack political discourse online [228, 229, 230]. While bots are certainly immune to unintended social influence, they may in turn influence the opinions of "real" social media users in different ways, for instance, by disseminating fake news [230] and low-credibility content [229], or artificially inflating support for a specific candidate in a political campaign [230]. In addition to artificial entities, such as bots, collective opinion formation might also be

strongly influenced by individuals who have committed to a particular political agenda, or who are specifically targeted by external communication sources [200, 202]. Specifically, their determination towards their opinion may render dyadic social influence asymmetrically, such that their peers are much more influence by them as vice versa.

To study similar phenomena in a linear model of social influence, let us return to the Taylor model [202], discussed in Ch. 3. Here, most agents are governed by assimilative forces, which leads to the alignment of their opinions. However, some agents are *stubborn* and committed to certain distinct opinion values, which may deviate from an existing consensus in the system. In this chapter we aim to answer the following questions: How will stubborn agents influence the overall opinion dynamics? What are the implications for the stationary opinion states? Is it possible to relate features of the underlying social influence network to the outcomes of the collective opinion dynamics?

To put our findings into the right context, we first briefly review some theoretical results on stubborn agents (Sec. 4.1) and highlight crucial differences to our approach. The subsequent sections are more technical. After discussing the adapted version of the Taylor model (Sec. 4.2), we present the Laplacian formalism, which aims to relate the properties of the emerging opinion dynamics to properties of the underlying social influence network (Sec. 4.3). This involves a novel complex network metric, which borrows its interpretation from the resistance distance, discussed in Ch. 2. Subsequently, we will apply the theory to two distinct settings, and on different influence networks, in order to shine light on some general implications of stubborn agents, within frameworks of continuous, assimilative opinion dynamics (Sec. 4.4 and 4.5). The results are discussed in Sec. 4.6.

4.1 Stubborn agents

Stubborn agents have been defined differently across the literature of opinion dynamics. While the Taylor model utilized in this chapter is formulated in a continuous opinion space, the main body of research investigates stubborn agents within models of discrete opinion dynamics [231, 232, 233, 234, 235, 236, 237, 238]

For instance, in the voter model (VM) stubborn agents were studied in Ref. [232], where they were modeled as agents with a bias towards one of the two opinions: in every time step a stubborn agent will change towards that opinion with probability γ. It was found that on one and two dimensional lattices, a single stubborn agent (also called zealot) influences an infinitely large system, and thus a global consensus emerges, where all agents hold the zealot's preferred opinion. In Ref. [231], a similar model was used to investigate the effects of multiple opposed (and completely inflexible) zealots. It was shown that already small groups of zealots are able to preclude a global consensus, or even the emergence of a large majority in the system. In the majority rule model stubborn agents were studied in [108]. It was found that, as in the case of the VM, a small group of inflexible agents, which do not change their opinions, profoundly impact the dynamics. Stubborn agents' opinions were found to dominate the final

opinion state, independent of the initial conditions. The effects of zealots and committed minorities was also investigated for the naming game, which is a simple model for opinion, cultural and language dynamics [237, 238]. Also here, small committed minorities of zeolots, which are not susceptible to be influenced by their peers, can rapidly reverse a prevailing majority [237], or restrict the opposed opinions to a tiny fraction of agents in the system [238].

In contrast to those previous studies, we investigate the effects of stubborn agents in a linear model of continuous opinion dynamics. The Taylor model is a fully deterministic approach, and it is formulated as a set of linear coupled differential equations, where the definition of stubborn agents differs from the ones in most discrete models. In the voter model stubborn agents' opinions are entirely fixed to a predefined value, or preferentially switch to certain, predefined opinions following a stochastic evolution. By contrast, in the Taylor model, stubborn agents are merely biased towards certain opinion values, while still being influenced by their peers in an assimilative and deterministic way.

4.2 Modified Taylor model

Deviating from the original definition of the Taylor model [see Eq. (3.15)], we will assume that a given communication source only influences a single agent i. We call those agents *stubborn agents*. We consider a coupled system of N agents, where each agent i (stubborn or not) holds a time-dependent opinion $x_i(t) \in \mathbb{R}$. Agent i is associated to a node i of an undirected[1] social influence network $\mathbf{W}$. The subset of stubborn agents is V_s and the number of stubborn agents is denoted as $N_s = |V_s|$.

The opinion of each regular agent is governed by the diffusive coupling between adjacent neighbors in the influence network, which directly corresponds to the Abelson model. Thus, we have

$$\dot{x}_i = -\sum_j W_{ij}\left(x_i - x_j\right), \quad i \notin V_s, \tag{4.1}$$

where $W_{ij} = W_{ji} \geq 0$ denotes the amount of influence that agent j exerts on agent i [cf. Eq. (3.6)]. The time-dependent opinions are written as x_i for brevity. As discussed in Ch. 3, Eq. (4.1) always reaches a perfect consensus for $t \to \infty$, if the underlying network is connected. In the following, we will focus on situations, where stubborn agents are present, and consensus will not necessarily emerge. In those cases, Eq. (4.1) is complemented by the equation governing the opinions of stubborn agents:

$$\dot{x}_i = -\sum_j W_{ij}\left(x_i - x_j\right) - \kappa[x_i - P_i(t)], \quad i \in V_s. \tag{4.2}$$

While stubborn agents take part in the opinion assimilation process, involving their connected

[1] This is in contrast to the original Abelson model, where the influence network $\mathbf{W}$ is not necessarily undirected, i.e., generally $W_{ij} \neq W_{ji}$.

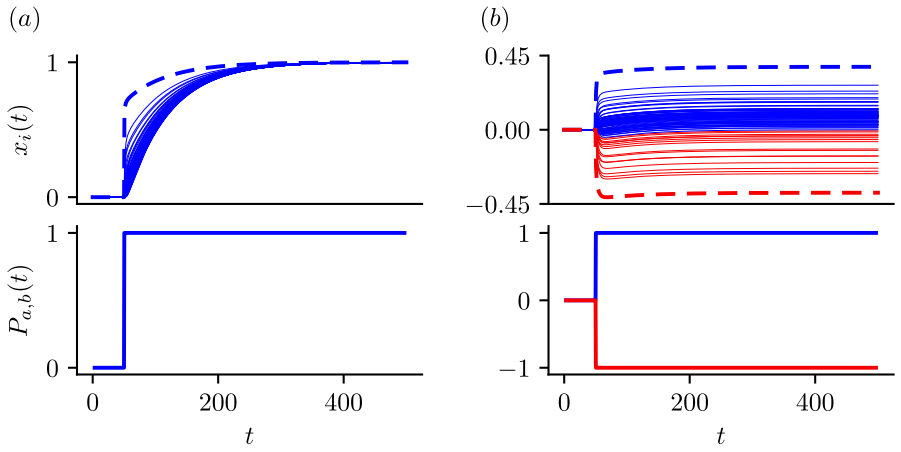

Figure 4.1: Opinion dynamics for one positively (a), and a pair of two antagonistically biased stubborn agents (b). After an initial consensus, a sudden change in the stubborn agents' biases, at $t = t_0$, induces the opinion dynamics on a WS network with $N = 100$ nodes. The opinion trajectories and the bias opinions are shown in the top and bottom panels, respectively. The colors emphasize the different trajectories of positively (blue) and negatively (red) biased agents (dashed lines) and the opinion associations of regular agents (thin solid lines) to one biased agent.

peers [sum in Eq. (4.2)], they are yet additionally influenced by individual opinion biases $P_i(t)$ [second term in Eq. (4.2)]. The rate of convergence towards their biases $P_i(t)$ is controlled by the parameter κ, which we refer to as *stubbornness*. Note that, for vanishing stubbornness ($\kappa = 0$), the model reduces to the original Abelson model, and stubborn agents behave like regular ones [see Eq. (4.1)].

To shed light on the role of stubborn agents in opinion assimilation models, we focus on two paradigmatic situations: (i) the transition from one consensus state to another induced by a single stubborn agent ($N_s = 1$), and (ii) the emergence of heterogeneous non-consensus states due to a pair of antagonistically biased agents a and b, with $P_a(t) = -P_b(t)$ ($N_s = 2$). In the first case, the goal is to quantify the persuasiveness of a single stubborn agent depending on their position in the social influence network: the opinion trajectories of the remaining (regular) agents will either follow closely the opinion of the stubborn agent (high persuasiveness), or their trajectories will be more spread out during this transient dynamics (low persuasiveness). In the second case, we aim to establish links between properties of the emerging non-consensus states, the positions of antagonistic stubborn agents, and the (statistical) properties of the underlying social networks.

Figure 4.1 illustrates the considered situations, where panels (a) and (b) correspond to the cases (i) and (ii), respectively. The top row shows the opinion evolution of all agents and the bottom row depicts the bias opinions of stubborn agents. The resulting opinion trajectories of the

positively ($x_a(t)$) and negatively ($x_b(t)$) biased agents are shown as blue and red dashed lines, respectively. Note that, the resulting opinion trajectories $x_a(t)$ and $x_b(t)$ do not correspond to the bias opinions of stubborn agents $P_a(t)$ and $P_b(t)$. The remaining agents' opinions are colored according to the stubborn agent they are associated to. This opinion association is defined in terms of opinion distance based on the final opinion state: a regular agent i is associated to the stubborn agent, to which they have the minimum opinion distance. In each of both considered cases, (i) and (ii), the population is initialized in a perfect consensus. Subsequently, at time $t = t_0$, the stubborn agents get biased towards opinions that are not aligned with the initial consensus. In Fig. 4.1(a), we depict the situation where a single stubborn agent $i = a$ gets biased towards $P_a(t > t_0) = 1$. After a transient period, during which the distribution of opinions may spread out substantially, all opinions x_i converge to the bias of agent a and a new (shifted) consensus is formed at $P_a = 1$. For two antagonistically biased agents $V_s = \{a, b\}$, with $P_a(t > t_0) = -P_b(t > t_0)$, the system does not reach a new consensus, as depicted in Fig. 4.1(b). Instead the regular agents' opinions are distributed over an interval of finite width, which is bounded by the final opinions of biased agents (dashed lines). Note that, for $N_s > 1$ the final opinion of a stubborn agent a, $x_a(t \to \infty)$, is generally not equal to their final bias opinion P_a [cf. Fig. 4.1(b)].

Broadly, our aim is to connect some features of the resulting collective dynamics to both the topology of the underlying influence network and the placement of stubborn agents therein. To this end, we utilize a Laplacian approach, which will be introduced below.

4.3 Laplacian formalism and modified resistance distances

The coupled sets of differential Eqs. (4.1) and (4.2) can be combined and written in a more compact vectorial form as

$$\dot{\mathbf{x}} = -(\mathbb{L} + \mathbf{K})\,\mathbf{x} + \kappa\mathbf{P}\,, \tag{4.3}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ denotes the vector of opinions. The matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$ is diagonal, with the element $K_{ii} = \kappa > 0$, if $i \in V_s$, and $K_{ii} = 0$, else. Similarly, the vector $\mathbf{P}$ contains the bias opinions of the stubborn agents, P_i, with vanishing components $P_i = 0$, for $i \notin V_s$. The purely diffusive part of the agents' interactions is captured in the Laplacian of the social influence network, whose definition we repeat for consistency (cf. Ch. 2):

$$\mathbb{L}_{ij} = \begin{cases} -W_{ij}\,, & i \neq j\,, \\ \sum_k W_{ik}\,, & i = j\,, \end{cases} \tag{4.4}$$

where $\mathbb{L}_{ij} = \mathbb{L}_{ji}$. The symmetry of $\mathbb{L}$ is a direct consequence of the assumed undirected influence network $\mathbf{W}$. Additionally, we define $\mathbb{L}^{(\kappa)} \equiv (\mathbb{L} + \mathbf{K})$, and refer to $\mathbb{L}^{(\kappa)}$ as the *modified Laplacian*. Since both $\mathbb{L}$ and $\mathbf{K}$ are symmetric matrices, also $\mathbb{L}^{(\kappa)}$ is symmetric.

The system of Eqs. (4.3) can be solved analytically using a spectral decomposition of $\mathbf{x}$,

$$\mathbf{x}(t) = \sum_{\alpha} c_{\alpha}(t)\, \mathbf{u}_{\alpha}^{(\kappa)}, \tag{4.5}$$

involving the orthonormal eigenvectors of the modified Laplacian, $\mathbf{u}_{\alpha}^{(\kappa)}$, where $c_{\alpha}(t)$ denotes the αth, time-dependent coefficient of this expansion. Plugging this spectral ansatz into the vectorial version of the Taylor model, Eq. (4.3), yields

$$\sum_{\alpha} \dot{c}_{\alpha}\, \mathbf{u}_{\alpha}^{(\kappa)} = -\sum_{\alpha} c_{\alpha}\, \mathbb{L}^{(\kappa)} \mathbf{u}_{\alpha}^{(\kappa)} + \kappa \mathbf{P} = -\sum_{\alpha} c_{\alpha}\, \lambda_{\alpha}^{(\kappa)} \mathbf{u}_{\alpha}^{(\kappa)} + \kappa \mathbf{P}, \tag{4.6}$$

where $\lambda_{\alpha}^{(\kappa)}$ is the αth eigenvalue of $\mathbb{L}^{(\kappa)}$. By multiplying Eq. (4.6) with an arbitrary eigenvector of $\mathbb{L}^{(\kappa)}$, $\mathbf{u}_{\beta}^{(\kappa)}$, and performing the summations the equation becomes

$$\dot{c}_{\alpha} = -\lambda_{\alpha}^{(\kappa)}\, c_{\alpha} + \kappa \mathbf{P} \cdot \mathbf{u}_{\alpha}^{(\kappa)}, \tag{4.7}$$

for $\alpha = 1, \ldots, N$, where we have used, that $\mathbf{u}_{\alpha}^{(\kappa)} \cdot \mathbf{u}_{\beta}^{(\kappa)} = \delta_{\alpha,\beta}$, with δ denoting the Kronecker delta. The general solution of Eq. (4.7) reads

$$c_{\alpha}(t) = c_{\alpha}(0) e^{-\lambda_{\alpha}^{(\kappa)} t} + \kappa e^{-\lambda_{\alpha}^{(\kappa)} t} \int_0^t e^{\lambda_{\alpha}^{(\kappa)} t'} \mathbf{P}(t') \cdot \mathbf{u}_{\alpha}^{(\kappa)} \mathrm{d}t'. \tag{4.8}$$

In combination with Eq. (4.5), it yields the time-evolution of the system for a given time-dependent bias vector $\mathbf{P}(t)$.

Certain aspects of the solution $\mathbf{x}(t)$ may be related to features of the influence network $\mathbf{W}$. The connection is established via the modified Laplacian and the developed framework involves a modified version of the resistance distance Ω_{ij} [163], which was originally defined for resistor networks, as discussed in Ch. 2. In contrast to Eq. (2.7), which defines the original resistance distance Ω_{ij} in terms of the regular Laplacian of the underlying network, we introduce the first-order *modified resistance distance* (MRD) as

$$\Omega_{ij}^{(\kappa,1)}(V_s) = \left[\mathbb{L}^{(\kappa)}\right]_{ii}^{-1} + \left[\mathbb{L}^{(\kappa)}\right]_{jj}^{-1} - \left[\mathbb{L}^{(\kappa)}\right]_{ij}^{-1} - \left[\mathbb{L}^{(\kappa)}\right]_{ji}^{-1}, \tag{4.9}$$

based on the modified Laplacian $\mathbb{L}^{(\kappa)}$. Just like $\mathbb{L}^{(\kappa)}$, the MRD depends on the specific set of stubborn agents V_s, and on the value of their stubbornness κ. Note that, in contrast to $\mathbb{L}$, the modified Laplacian $\mathbb{L}^{(\kappa)}$ is not a singular matrix [2]. Accordingly, $\mathbb{L}^{(\kappa)}$ is invertible and modified resistance distances are simply defined in terms of its inverse $\left[\mathbb{L}^{(\kappa)}\right]^{-1}$, rather than its pseudo-inverse, cf. Eq. (2.7). Following [161], we may generalize $\Omega_{ij}^{(\kappa,1)}$ to the pth power

[2] In App. A, we show that all eigenvalues of $\mathbb{L}^{(\kappa)}$ are positive, i.e. , we find that $\lambda_{\alpha}^{(\kappa)} > 0$.

using the eigenvectors and the eigenvalues of $\mathbb{L}^{(\kappa)}$. This yields

$$\Omega_{ij}^{(\kappa,p)}(V_s) = \left[\mathbb{L}^{(\kappa)}\right]_{ii}^{-p} + \left[\mathbb{L}^{(\kappa)}\right]_{jj}^{-p} - \left[\mathbb{L}^{(\kappa)}\right]_{ij}^{-p} - \left[\mathbb{L}^{(\kappa)}\right]_{ji}^{-p} \tag{4.10}$$

$$= \sum_{\alpha} \frac{\left(u_{\alpha,i}^{(\kappa)} - u_{\alpha,j}^{(\kappa)}\right)^2}{\left(\lambda_{\alpha}^{(\kappa)}\right)^p}. \tag{4.11}$$

By means of this general-order MRD, between two nodes i and j, we may also define the associated closeness centralities

$$C_p(i, V_s) = \left[N^{-1} \sum_j \Omega_{ij}^{(\kappa,p)}(V_s) \right]^{-1}. \tag{4.12}$$

In the spirit of traditional centrality measures [159, 239], $C_p(i, V_s)$ quantifies the average MRD (of order p) from node i to all other nodes j in the network $\mathbf{W}$, given a set of stubborn agents V_s. Thus, for large (small) values of $C_p(i, V_s)$ node i is central (peripheral), with respect to $\Omega_{ij}^{(\kappa,p)}$. It is important to emphasize that the modified resistance distance and all quantities derived from it depend on the specific set of stubborn agents and need to be re-evaluated if V_s changes; although the influence network $\mathbf{W}$ does not change.

4.4 Consensus change

In this section, we consider the consensus change induced by a single stubborn a. The biased agent drives the system from an initial consensus state to a new one, given by the final bias P_a. Inspired by previous works on leader–follower systems [240, 241], we aim to quantify how closely, or coherently, the remaining agents follow the opinion trajectory of the stubborn agent a. Specifically, we consider a setup in which initially all agents' opinions, and the bias of the stubborn agent a, form a consensus, i.e., we assume that $P_a(0) = x_i(0)\ \forall i$. Subsequently, at time $t = t_0$, the bias value $P_a(t)$ abruptly changes towards a value different from the initial consensus, modeled in terms of a Heaviside step function: $P_a(t) = P\theta(t - t_0)$, with $\theta(x) = 0$, for $x \leq 0$, and $\theta(x) = 1$ otherwise. Without loss of generality, we additionally assume a positive final bias of the single stubborn agent, i.e., $P(t \to \infty) = P > 0$. Figure 4.1(a) illustrates the time-dependent bias opinion $P_a(t)$ (bottom) and the induced dynamics (top), where the opinion trajectory of the single stubborn agent, $x_a(t)$, is depicted as thick dashed line and the regular agents' opinions are shown as thin solid lines.

To systematically quantify the transient opinion spread and by that the persuasiveness of stubborn agent a, we define the coherence measure

$$\mathscr{C}(a) = \sum_i \int_0^{\infty} |x_a(t) - x_i(t)|\, \mathrm{d}t, \tag{4.13}$$

given as the sum of all integrated opinion distances between each agent i and the stubborn

agent a, during the transition from the consensus at $\mathbf{x}(t=0)=0$, to the new consensus at $\mathbf{x}(t\to\infty)=P$. Note, that by definition, the coherence of the induced transition is high (low) for small (large) values of $\mathscr{C}$.

Within the spectral decomposition approach, $\mathscr{C}(a)$ can be expressed solely in terms of quantities derived from $\mathbb{L}^{(\kappa)}$. Inserting Eq. (4.8) into Eq. (4.5), yields the general solution of $x_i(t)$ in the case of a single stubborn agent. While for $t<t_0$ the system prevails in a consensus at $x_i(t)=0$, the solution for $t\geq t_0$, describes its transient behavior, and reads

$$x_i(t)=\kappa P\sum_\alpha \frac{u_{\alpha,a}^{(\kappa)}}{\lambda_\alpha^{(\kappa)}}\left(1-e^{-\lambda_\alpha^{(\kappa)}(t-t_0)}\right)u_{\alpha,i}^{(\kappa)}. \tag{4.14}$$

In the limit of $t\to\infty$, Eq. (4.14) becomes

$$x_i(t\to\infty)=\kappa P\sum_\alpha \frac{u_{\alpha,a}^{(\kappa)}}{\lambda_\alpha^{(\kappa)}}u_{\alpha,i}^{(\kappa)}=P, \tag{4.15}$$

where it was used that $\sum_\alpha u_{\alpha,a}^{(\kappa)}u_{\alpha,j}^{(\kappa)}/\lambda_\alpha^{(\kappa)}=1/\kappa$, as derived in App. A. Equation (4.15) ensures the formation of a new consensus at $x_i(t\to\infty)=P$ as final state of the system. Note that this consensus emerges independent of the network structure, the level of stubbornness κ, or the position of the stubborn agent; only the transient dynamics toward the new consensus is influenced by these latter factors.

To obtain a closed form expression for the coherence measure $\mathscr{C}(a)$, Eq. (4.14) is plugged into Eq. (4.13). Performing the time integration yields

$$\mathscr{C}(a)=\kappa P\sum_i\sum_\alpha \frac{u_{\alpha,a}^{(\kappa)}u_{\alpha,i}^{(\kappa)}-\left(u_{\alpha,a}^{(\kappa)}\right)^2}{\lambda_\alpha^{(\kappa)}}, \tag{4.16}$$

where $\mathscr{C}(a)$ is solely defined in terms of the eigenvectors and eigenvalues of $\mathbb{L}^{(\kappa)}$. Note that, the coherence measure presented in Eq. (4.16) can be generalized for the case of multiple stubborn agents, with the same bias, by performing an additional sum over all biased agents $a\in V_s$.

According to Eq. (4.16), the value of $\mathscr{C}(a)$ depends on the node index $i=a$, on which the single biased agent is placed. This is illustrated by the results shown in Fig. 4.2(a) for Watts-Strogatz networks (WS) (top) and networks generated by the stochastic block model (SBM). Each node is colored according to the resulting value of $\mathscr{C}(a)$, and the brightness increases with the value of $\mathscr{C}(a)$. Accordingly, if a biased agent is placed on a bright node, regular agents do not follow their trajectory closely. Instead, they are rather widely spread, similar to the case depicted in the top panel of Fig. 4.1(a).

A closer look at Fig. 4.2(a) furthermore suggests that, for both networks, the distribution of $\mathscr{C}(a)$ is related to the degree k_a of the stubborn agent's node, with low (high) degree nodes

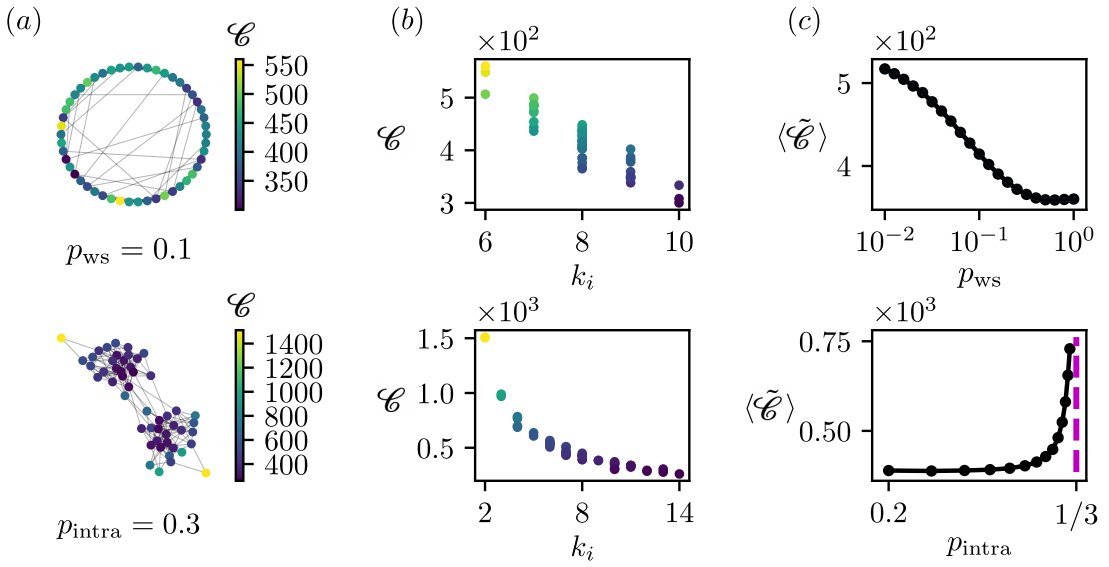

Figure 4.2: Coherence of consensus change induced by a single stubborn agent on WS (top panels) and SBM (bottom panels). Panel (a): single realization of a WS network (top) and a SBM network (bottom), where each node a is colored according to the resulting coherence value $\mathscr{C}(a)$. Panel (b): relation between $\mathscr{C}(a)$ and the degree k_a of node a. Panel (c): average opinion coherence $\langle\tilde{\mathscr{C}}\rangle$ as a function of the p_{ws} and p_{intra} for WS (top) and SBM (bottom) networks, respectively. All networks are comprised of $N = 50$ nodes and (on average) $N_E = 200$ edges, thus we set $k_{\mathrm{ws}} = 8$ (in the WS model) and implemented the relation Eq. (A.34) (for the SBM).

resulting in large (small) values of $\mathscr{C}(a)$. This is indeed the case, as shown in Fig. 4.2(b), where the opinion coherence of $\mathscr{C}(a)$ is plotted against the corresponding degree k_a, revealing a clear negative correlation. However, the degree k_a cannot be utilized as an unambiguous predictor for $\mathscr{C}(a)$, as for some nodes of different degrees, the corresponding coherence values overlap. This effect becomes less pronounced for increased edge densities, as demonstrated in Fig. A.1(a) of App. A. Within a given degree class the values of $\mathscr{C}$ also may vary substantially, as shown for the WS network [top panel of Fig. 4.2(b)].

From this variability of $\mathscr{C}(a)$ across a given network it follows that the single node perspective is not sufficient to systematically characterize opinion coherence on a network level. Instead all nodes should be included in the analysis. Therefore, we define

$$\tilde{\mathscr{C}} = N^{-1} \sum_i \mathscr{C}(i)\,, \tag{4.17}$$

which defines the average opinion coherence computed over all nodes in a given network. In Fig. 4.2(c), $\tilde{\mathscr{C}}$, is shown as a function of p_{ws} and p_{intra}, for WS (top) and SBM networks (bottom), respectively. As discussed in Ch. 2, the parameter p_{ws} denotes the re-wiring probability of WS networks and p_{intra} defines the probability for connections between two nodes within the same community in SBM networks. To fix the total number of edges in SBM networks, we implement a relation between p_{intra} and p_{inter}, where the latter is the probability for links between nodes in different communities (see App. A for details). Both WS and SBM networks are random networks, and thus subject to disorder. Therefore, we depict $\tilde{\mathscr{C}}$ not for a single network, but as an average over 5000 distinct networks generated for the same values of p_{ws} (WS) and p_{intra} (SBM), respectively, denoted as $\langle\tilde{\mathscr{C}}\rangle$.

For WS networks, we observe a monotonic decrease of $\langle\tilde{\mathscr{C}}\rangle$ for increasing re-wiring probabilities [see top panel of Fig. 4.2(c)]. For SBM networks, low and intermediate values of p_{intra} yield small values for the opinion coherence, which remain nearly constant, as shown in the bottom panel of Fig. 4.2(c). As p_{intra} further increases, $\langle\tilde{\mathscr{C}}\rangle$ diverges. By construction, the divergence is precisely located at $p_{\text{intra}} = 1/3$, for which the network disintegrates in two parts (see App. A for more details). In this case, a biased agent in the first community of the network cannot influence agents in the second community. Accordingly, the integral in Eq. (4.13) diverges, as half of the agents remain at the initial consensus value $x = 0$.

4.5 Heterogeneous opinion states

A single stubborn agent cannot induce a persistent non-consensus state in the limit of $t \to \infty$: they may only shift a prevailing consensus to a new value. By contrast, heterogeneous opinion states may emerge and persist if multiple stubborn agents with different biases are introduced. In the following, we will investigate the properties of such non-consensus states within a minimal setting of two stubborn agents.

More specifically, we consider a situation of two antagonistically biased agents $V_s = \{a, b\}$. For

simplicity, the agents are assumed to have perfectly balanced biases, i.e., $P_a(t) = P\Theta(t - t_0) = -P_b(t)$. Such a case is illustrated in Fig. 4.1(b), where indeed a stable heterogeneous opinion state emerges. The agents' opinions do not form a consensus, but are distributed over a finite interval. As in the previous case of a single stubborn agent, the system is initialized in a state of perfect consensus at $x_i(0) = 0$. Then, at $t = t_0$, the stubborn agents a and b develop opposite biases towards a positive ($P_a > 0$) and a negative ($P_b < 0$) opinion with magnitude P, respectively. In this case, the time-evolution of agent i can be derived similarly to Eq. (4.14). For $t \geq t_0$, the opinion of agent i evolves as

$$x_i(t) = \kappa P \sum_{\alpha} \frac{u^{(\kappa)}_{\alpha,a} - u^{(\kappa)}_{\alpha,b}}{\lambda^{(\kappa)}_{\alpha}} \left(1 - e^{-\lambda^{(\kappa)}_{\alpha}(t-t_0)}\right) u^{(\kappa)}_{\alpha,i}, \quad \text{for} \quad i = 1, \ldots, N, \tag{4.18}$$

which yields,

$$x_i(t \to \infty) = \kappa P \sum_{\alpha} \frac{u^{(\kappa)}_{\alpha,a} - u^{(\kappa)}_{\alpha,b}}{\lambda^{(\kappa)}_{\alpha}} u^{(\kappa)}_{\alpha,i}, \tag{4.19}$$

in the limit of $t \to \infty$. Note that, $x_i(t \to \infty)$, as defined in Eq. (4.19), merely changes its sign upon an interchange of indices a and b, which follows from $\sum_\alpha {u^{(\kappa)}_{\alpha,a}}^2 / \lambda^{(\kappa)}_{\alpha} = \sum_\alpha {u^{(\kappa)}_{\alpha,b}}^2 \lambda^{(\kappa)}_{\alpha}$, as derived in App. A. Thus, the final opinions of biased agents have the same magnitude, i.e., it holds that $|x_a(t \to \infty)| = |x_b(t \to \infty)|$. Starting from Eq. (4.19), we will analytically quantify some properties of the emerging non-consensus states, which can be related to features of the underlying influence networks, by leveraging the concept of modified resistance distances.

Opinion association

In the case of two opposed biased agents it is natural to ask, which of the two agents maximizes their influence across the network. In particular, we may ask with respect to the final distribution of opinions: which stubborn agent influences the majority, or specifically relevant agents? Similar questions evolve around influence maximization in social networks, where it is aimed to exploit the positions of an individual's network position for purposes such as viral marketing [242] or the diffusion of innovation [243]. Here, we investigate the problem from a slightly different angle. We consider two competing entities, the stubborn agents a and b, which may influence the remaining regular agents in the network. This "competition" is operationalized in terms of an opinion association framework. We define that a regular agent i is associated to the one stubborn agent to which they have the smaller opinion distance to in the final opinion state. Therefore, it is their opinion stance, i.e., the sign of their final opinion, $\mathrm{sgn}(x_i(t \to \infty))$, that is of interest, rather than their conviction $|x_i(t \to \infty)|$. This view is motivated by the binary character of many real–world decisions, such as presidential elections, or political referendums, where two opposing parties try to convince the majority of individuals in a given population into a certain direction.

The introduced Laplacian framework yields an efficient solution to this association problem,

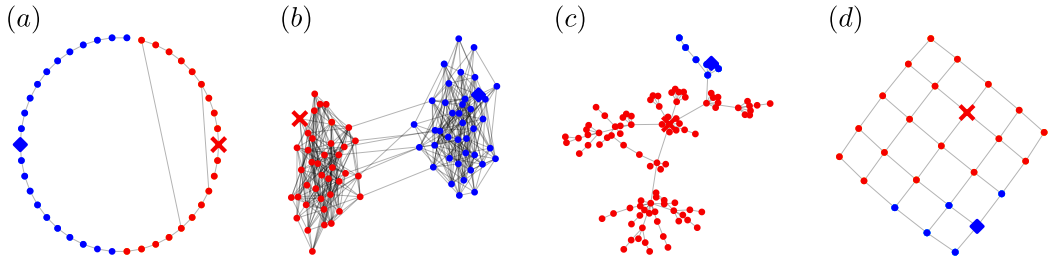

Figure 4.3: Opinion associations resulting from two opposed stubborn agents on four different networks: a ring network with two randomized edges (a), a SBM network (b), a BA network (c), and a two-dimensional lattice (d). The stubborn agents' positions are depicted as a blue rhombus (positively biased agent) and a red cross (negatively biased agent). The remaining nodes are colored according to their opinion associations, defined in Eq. (4.20).

in terms of modified resistance distances. The opinion of each regular agent i is generally shifted towards either a positive or a negative value, which generally yields a shorter opinion distance to the positively (negatively) biased agent[3]. Using the definition of the first-order MRD, Eq. (4.19) can be compactly written as

$$x_i^{\infty} = \frac{\kappa P}{2}\left[\Omega_{bi}^{(\kappa,1)}(\{a,b\}) - \Omega_{ai}^{(\kappa,1)}(\{a,b\})\right], \tag{4.20}$$

where x_i^{∞} denotes the final opinion of agent i (see App. A for details on the derivation). Notably, Eq. (4.20) shows how to reformulate the problem of opinion associations, which was originally defined in the space of opinions, in terms of modified resistance distances. To determine the opinion association of agent i, one can simply compute the sign of Eq. (4.20). In some highly symmetric networks, an agent i may have identical opinion distances to both biased agents. In the space of MRDs, this special case translates into $x_i^{\infty} = \Omega_{bi}^{(\kappa,1)} - \Omega_{ai}^{(\kappa,1)} = 0$.

The applicability of our framework to the opinion association problem is demonstrated in Fig. 4.3, where we depict the evaluation of Eq. (4.20) on four different networks. The positively (negatively) biased stubborn agent is depicted as blue rhombus (red cross), and the remaining agents are shown as dots, colored according to their final opinion associations. In Fig. 4.3(a), the two stubborn agents are placed on opposite sides of a ring network, where two edges have been rewired according to the WS model. The rewiring has only slightly changed the MRDs between each biased agent and their corresponding opposite sides of the network, as no direct shortcuts were introduced. Thus, the opinion association of agents is roughly split into two half circles of approximately equal size, where agents are associated to the one stubborn agent which is placed on their side. For SBM networks, a high value of p_{intra} leads to two pronounced communities. If we place one biased agent in each of these communities, this leads most

[3] This follows from Eq. (4.19), which shows that the final convictions of stubborn agents are equal, i.e., $x_a(t \to \infty) = -x_b(t \to \infty)$.

probably to a situation depicted in Fig. 4.3(b). Due to the high density of intra-community connections the MRDs within each community are drastically reduced, such that a stubborn agent will attract all remaining agents within that community. Figure 4.3(c) shows the case of a BA network, where the positively biased agent is placed on the node with the highest degree, and the negatively biased agent resides on a node of degree $k = 1$. While the positive agent attracts the majority of the agents, only a few peripheral agents are associated to the negatively biased agent. This is due to the high degree of the positive agent, which leads to a strong reduction of their MRDs to most other agents in the network, and thus results in a high centrality value, as defined in Eq. (4.12). Finally, we briefly discuss the case of a two dimensional lattice, shown in Fig. 4.3(d). Here, a similar picture arises, as observed in the BA network. The positively stubborn agent, which is placed at a central position in the lattice, has smaller MRDs to most other agents and therefore attracts more regular agents than the one placed at the periphery of the lattice.

Properties of heterogeneous opinion states

Going beyond the binary character of opinion associations, it is desirable to examine more closely the properties of the emerging non-consensus states. As for opinion associations, expressing those properties in terms of MRDs allows to draw connections between the emerging opinion dynamics and some structural features of the social influence networks. The following discussion will focus on three aspects of heterogeneous opinion states: (i) the largest opinion distance in the population, (ii) the mean of the final opinion distribution, and (iii) its variance. While the properties (i) and (iii) capture the degree of opinion heterogeneity, (ii) quantifies how strongly one of the stubborn agents is favored over the other.

For two opposed biased agents, $V_s = \{a, b\}$, with $P_a(t) = P\Theta(t - t_0) = -P_b(t)$, the largest opinion distance is given as the difference of stubborn agents' final opinions, denoted as $D_{\max} = x_a^\infty - x_b^\infty$. From Eq. (4.19), it follows that $D_{\max}$ can be expressed as

$$
\begin{aligned}
D_{\max} &= x_a^\infty - x_b^\infty && \text{(4.21a)}\\
&= \kappa P \left(\sum_\alpha \frac{\left(u_{\alpha,a}^{(\kappa)} - u_{\alpha,b}^{(\kappa)}\right) u_{\alpha,a}^{(\kappa)}}{\lambda_\alpha^{(\kappa)}} - \frac{\left(u_{\alpha,a}^{(\kappa)} - u_{\alpha,b}^{(\kappa)}\right) u_{\alpha,b}^{(\kappa)}}{\lambda_\alpha^{(\kappa)}} \right) && \text{(4.21b)}\\
&= \kappa P \sum_\alpha \frac{\left(u_{\alpha,a}^{(\kappa)} - u_{\alpha,b}^{(\kappa)}\right)^2}{\lambda_\alpha^{(\kappa)}} && \text{(4.21c)}\\
&= \kappa P \, \Omega_{ab}^{(\kappa,1)}, && \text{(4.21d)}
\end{aligned}
$$

involving the MRD between both biased agents a and b, $\Omega_{ab}^{(\kappa,1)}$, where in the last step we used Eq. (4.10). The maximum opinion distance is bounded and stubborn agents do generally not reach their individual opinion biases of magnitude P. As shown in App. A, we have $D_{\max} \leq 2P$ where the equality holds in the limit of $\kappa \to \infty$, i.e., in the case of infinitely stubborn agents.

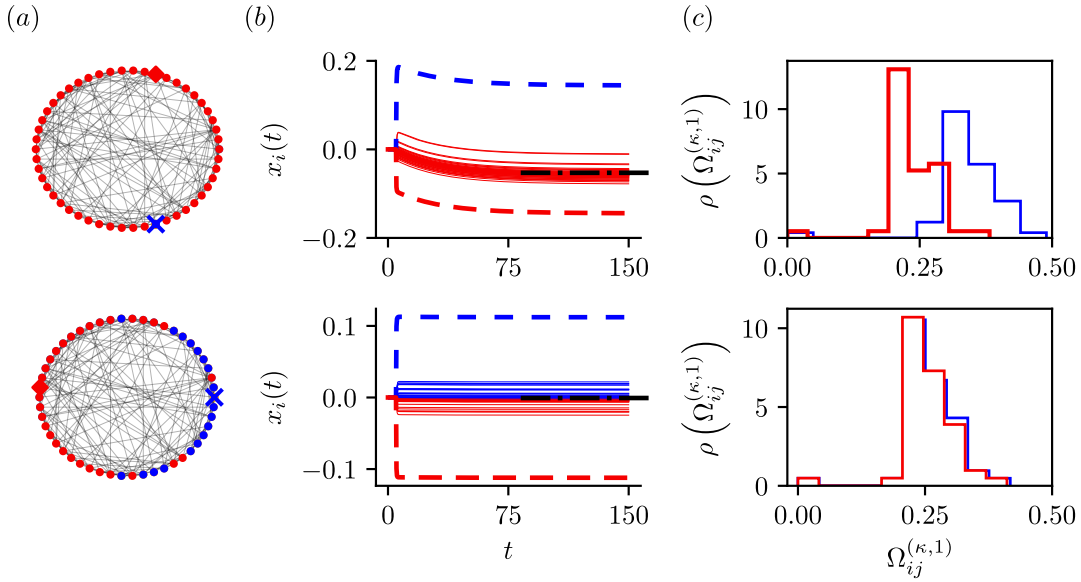

Figure 4.4: Position-dependent effects of two antagonistically biased stubborn agents on a fixed influence network. Panel (*a*) shows the resulting opinion associations for two different sets of stubborn agents. The red rhombus (blue cross) depicts the position of the negatively (positively) biased agent. Panel (*b*) shows the transient opinion dynamics where the thick dashed lines correspond the opinion trajectories of the stubborn agents, and thin lines depict the opinion evolution of regular agents, colored according to their opinion associations. Panel (*c*) shows the distributions of MRDs, $\Omega_{ij}^{(\kappa,1)}$, of stubborn agent i to all remaining agents j, colored according to their positive (blue) and negative (red) biases. The results are shown for a WS network of $N = 50$ nodes.

The mean $\mu_x = N^{-1}\sum_i x_i$ and the variance $\sigma_x^2 = N^{-1}\sum_i (x_i - \mu_i)^2$ of the final opinion distribution can also be expressed in terms of MRDs. This yields the following expressions

$$\mu_x(\{a,b\}) = \frac{\kappa P}{2}\left[C_1^{-1}(b) - C_1^{-1}(a)\right] \tag{4.22}$$

and

$$\sigma_x^2(\{a,b\}) = \left(\frac{\kappa P}{2}\right)^2 \left(\frac{4\Omega_{ab}^{(\kappa,2)}}{N} - \left[C_1^{-1}(b) - C_1^{-1}(a)\right]^2\right); \tag{4.23}$$

which are derived in App. A. Note that, second-order MRDs can be expressed in terms of first-order ones as $\Omega_{ab}^{(\kappa,2)} = (1/4)\sum_i \left(\Omega_{bi}^{(\kappa,1)} - \Omega_{ai}^{(\kappa,1)}\right)^2$, such that the quantities $D_{\max}$, μ_x, and σ_x^2 can all be formulated solely in terms of $\Omega_{ij}^{(\kappa,1)}$.

As suggested by Eqs. (4.21)–(4.23), important characteristics of the final opinion states are determined by both the topology of the underlying influence network and the placement

of the biased agents. To demonstrate the latter, in Fig. 4.4, we show the resulting opinion dynamics on a fixed network topology for different positions of the stubborn agents. First, we note that the associations of opinions (colored according to the associated biased agent) strongly differ between both sets of biased agents, where the positive and negative biased agents are depicted as blue cross and red rhombus, respectively [see Fig. 4.4(a)]. While the positively biased agent (blue cross) does not attract a single agent for the first set V_s (top panel), almost half of the agents show a positive final opinion for the second one (bottom panel). The opinion associations are reflected in the transient dynamics $x_i(t)$ towards the final state, shown in Fig. 4.4(b). The opinions of stubborn agents are depicted as dashed blue ($i = a$) and dashed red ($i = b$) lines, and thin solid lines show the opinion trajectories of the remaining agents, colored according to their opinion associations. The black dashed dotted line shows the final mean opinion μ_x , which is close to zero in the balanced case (bottom panel), and clearly shifted to a negative value in the top panel of Fig. 4.4(b). Interestingly, these features of the final opinion states can be understood in terms of the distributions of MRDs, depicted in Fig. 4.4(c). Here, $\rho\left(\Omega_{ij}^{(\kappa,1)}\right)$ denotes the normalized histogram of MRDs of a stubborn agent i to all other agents j in the population, and the colors of the histograms correspond to the positively (blue) and negatively (red) biased agents. For the more balanced state, the distributions essentially overlap (bottom panel). Instead, as shown in the top panel, the red histogram is strongly shifted to lower values for the first set of stubborn agents. As a consequence, a strongly unbalanced situation arises, where all regular agents are associated to the positively biased agent.

The strong dependence of the resulting opinion distribution on the specific set of biased agents suggest that, Eqs. (4.21)–(4.23) will only be useful if V_s is known. To account for cases in which the exact positions of both biased agents are unknown, and to establish measures of $D_{\max}$, μ_x, and σ_x^2 on a network level, we define

$$\tilde{D}_{\max} = \left(\frac{N(N-1)}{2}\right)^{-1} \sum_{a<b} D_{\max}(\{a,b\}) \tag{4.24a}$$

$$\tilde{\mu}_x = \left(\frac{N(N-1)}{2}\right)^{-1} \sum_{a<b} |\mu_x(\{a,b\})| \tag{4.24b}$$

$$\tilde{\sigma_x^2} = \left(\frac{N(N-1)}{2}\right)^{-1} \sum_{a<b} \sigma_x^2(\{a,b\})\,. \tag{4.24c}$$

Note that for $\tilde{\mu}_x$ we consider the absolute value of μ_x, as we are only interested in the mean deviation from $x = 0$ irrespective of its sign.

In Fig. 4.5, those global network measures are evaluated on WS and SBM networks. As in the former case of opinion coherence, we show $\tilde{D}_{\max}$, $\tilde{\mu}_x$ and $\tilde{\sigma_x^2}$ as functions of p_{ws} (WS) and p_{intra} (SBM) and averaged ($\langle\cdot\rangle$) over 5000 instances of random WS and SBM networks. First, we notice that both $\langle\tilde{D}_{\max}\rangle$ and $\langle\tilde{\sigma_x^2}\rangle$ behave very similarly to the coherence measure $\langle\tilde{\mathscr{C}}\rangle$. For the WS model an increased rewiring probability leads to a decrease in both quantities [see bottom and top panel of Fig. 4.5(a)]: $D_{\max}$ is proportional to the MRDs between both biased agents

(Eq. (4.21d)), and such distances are strongly reduced for increasing rewiring, as random shortcuts are added to the networks. This effect is particularly pronounced in the range of small values of $p_{\mathrm{ws}} \simeq 0.1$. Larger values of p_{ws} yield the saturation of $\langle \tilde{D}_{\mathrm{max}} \rangle$ to a value, which is expected for a random network with the same number of nodes and edges. Arguments along these lines may also be used to explain the decrease of $\langle \tilde{\sigma}_x^2 \rangle$ for increasing p_{ws}. As suggested by Eq. (4.23), the opinion variance can be expressed as the difference between a term that is proportional to $\Omega_{ab}^{(\kappa,2)}$ (the second-order MRD between the two biased agents a and b), and the squared opinion mean μ_x^2. Thus, $\langle \tilde{\sigma}_x^2 \rangle$ decreases for increasing p_{ws}, since $\Omega_{ab}^{(\kappa,2)}$ is on average reduced between agents due to re-wiring, while $\langle \mu_x \rangle$ increases, as will be discussed next. As a consequence of the perfectly symmetric structure of WS networks for $p_{\mathrm{ws}} = 0$ (ring lattice), the centrality values [Eq. (4.12)] of both stubborn agents are generally equal, i.e. $C_1(a) = C_1(b)$, yielding $\mu_x = 0$ (Eq. (4.22)). The introduction of randomness breaks this symmetry, leading to higher values of $\langle \tilde{\mu}_x \rangle$ for $p_{\mathrm{ws}} > 0$. Note that, although $\langle \mu_x \rangle = 0$ for $p_{\mathrm{ws}} = 0$, this does not imply the establishment of a perfect consensus. Instead, the population will split into two equally sized groups of nodes, with each half being associated to one of the stubborn agents. This is indicated by the finite values of $\langle \tilde{D}_{\mathrm{max}} \rangle$ and $\langle \tilde{\sigma^2}_x \rangle$ for $p_{\mathrm{ws}} = 0$.

In Fig. 4.5(b), we show the results for SBM networks. Here, for increasing values of the intra–community connection probability (p_{intra}) the values of both $\langle \tilde{D}_{\mathrm{max}} \rangle$ and $\langle \tilde{\sigma^2}_x \rangle$ increase monotonically, which was similarly observed for $\langle \tilde{\mathscr{C}} \rangle$ [see Fig. 4.2(c)]. With increasing p_{intra} the density of edges between the two communities is decreased. For the assumed case of equally sized communities there are more sets of stubborn agents (V_s), where the stubborn agents are found in distinct communities, rather than in the same one. The decreasing connectivity between both communities, which leads to a strongly increased MRD between them, therefore leads to the increase of $\langle \tilde{D}_{\mathrm{max}} \rangle$ and ultimately its divergence. Analogously, the second-order modified resistance distance between a random pair of stubborn agents ($\Omega_{ab}^{(\kappa,2)}$), is on average increased, such that $\langle \tilde{\sigma}_x^2 \rangle$ increases with p_{intra} . As is the case for WS networks, $\langle \tilde{\mu}_x \rangle$ also deviates from the behavior of both $\langle \tilde{D}_{\mathrm{max}} \rangle$ and $\langle \tilde{\sigma}_x^2 \rangle$ on SBM networks. After an initial increase to its maximum value, it strongly decreases, as p_{intra} approaches the value of 1/3. For $p_{\mathrm{intra}} = 1/3$ only nodes within the same community are connected, leaving the network disintegrated in two equally sized parts. The decrease of $\langle \tilde{\mu}_x \rangle$, while approaching the limit of $p_{\mathrm{intra}} = 1/3$ can be explained as follows. For a pronounced community structure most sets of stubborn agents V_s lead to a rather balanced situation, as the one depicted in Fig. 4.3(b): all nodes of one community are associated to the stubborn agent placed therein. Also the cases where both stubborn agents reside in the same community are expected to result in smaller values of $\tilde{\mu}_x$: with increasing p_{intra} the bounds of the opinion distribution decrease ($\tilde{D}_{\mathrm{max}}$), which in turn limits the absolute value of the final mean opinion.

The total number of agents N, edges N_E, and the value of the biased agents' stubbornness κ were held constant for the results shown in Fig. 4.5. In the following, we will investigate their influence on $\langle \tilde{\mu}_x \rangle$ and $\langle \tilde{D}_{\mathrm{max}} \rangle$ for WS networks. In contrast to the bias magnitude P, which multiplies Eq. (4.21) and (4.22) as a prefactor, the stubbornness κ also enters the definition of the modified Laplacian $\mathbb{L}^{(\kappa)}$ and may therefore alter the opinion formation process in a

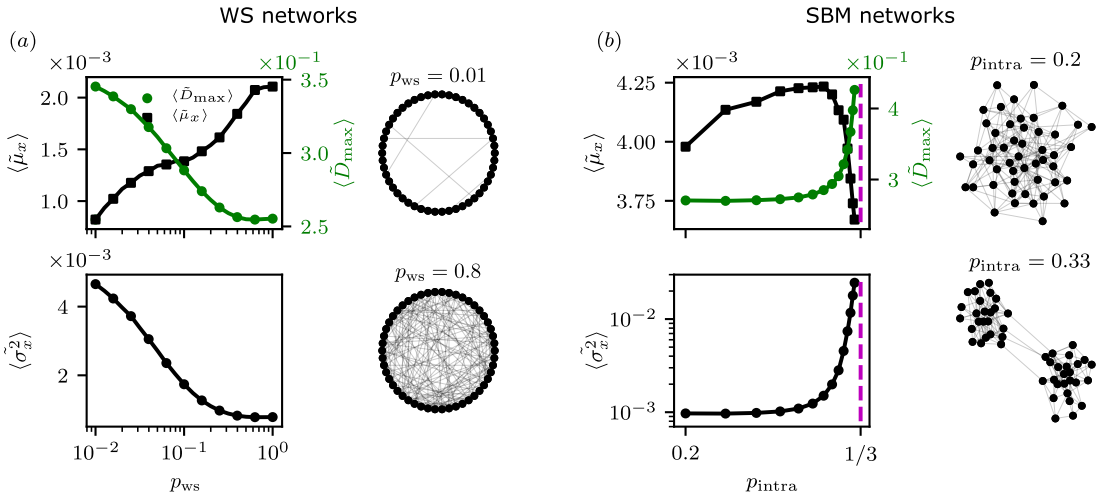

Figure 4.5: Derived measures for heterogeneous opinion states $\tilde{D}_{\max}$, $\tilde{\mu}_x$ and $\tilde{\sigma}_x^2$ on WS (*a*) and SBM (*b*) networks as functions of p_{ws} and p_{intra}, respectively. For both network models the number of nodes and edges is (on average) fixed to $N = 50$ and $N_E = 200$. The quantities $\langle \tilde{D}_{\max} \rangle$, $\langle \tilde{\mu}_x \rangle$ and $\langle \tilde{\sigma}_x^2 \rangle$ are computed as averages over 5000 WS networks.

non-linear way. The results are shown in Fig. 4.6. First, in panel (a) the total number of nodes $N \in [30, 40, 50]$ and edges $N_E \in [120, 160, 200]$ are varied, fixing the average degree to $\langle k \rangle = 8$. The black cross symbols correspond to the configuration discussed in Fig. 4.5(a). For large p_{ws}, decreasing numbers of nodes and edges yield increasing values of $\tilde{\mu}_x$, an effect which is inverted for $p_{\mathrm{ws}} < 0.1$. On the contrary, $\langle \tilde{D}_{\max} \rangle$ is reduced for decreasing system sizes across the whole p_{ws}–range. This can be attributed to the overall increased MRDs, between a random pair of nodes, as the system gets larger. In panel (b) the number of agents is held constant ($N = 50$), while the number of edges is varied. Decreasing N_E promotes larger values of the mean average opinion, and alters the shape of $\langle \tilde{\mu}_x \rangle$, with two local maxima arising for $N_E = 150$. The average maximum opinion distance is affected similarly: adding edges to the network leads to smaller values of $\langle \tilde{D}_{\max} \rangle$. Due to the increased edge density (N is held constant) the MRD between a random pair of nodes is on average decreased, which yields, according to Eq. (4.21), smaller values of $D_{\max}$. The reduction of $\langle \tilde{D}_{\max} \rangle$ as a function of N_E is directly related to the decrease of $\langle \tilde{\mu}_x \rangle$ for higher edge densities, shown in the top panel of Fig. 4.6(b). Intuitively, while more and more edges are added, the structure gradually turns into a complete graph. This yields vanishing differences of centralities (Eq. (4.12)), between any chosen pair of biased agents [cf. Eq. (4.22)]. The effects of different levels of stubbornness are depicted in Fig. 4.6(c). While the overall shape of $\tilde{\mu}_x$ is approximately retained, its values increase for higher stubbornness κ. A similar behavior is observed for $\tilde{D}_{\max}$, which approaches the value of $\tilde{D}_{\max} = 2P$ in the limit of $\kappa \to \infty$, as depicted by the pink dashed line and derived in App. A.

Finally, we briefly discuss how the structural randomization of an empirical friendship network

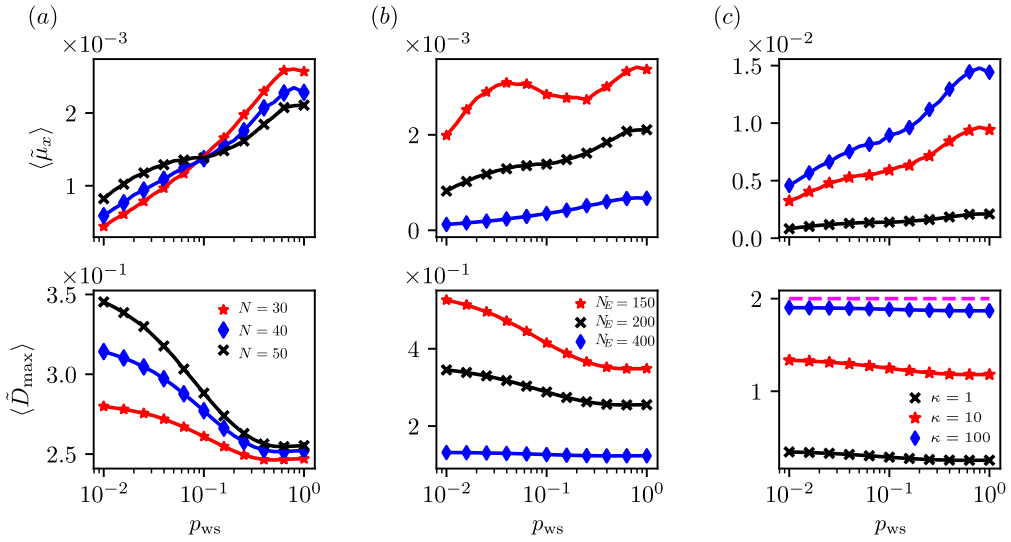

Figure 4.6: Mean opinion and maximum opinion distance for different system parameters. Panels (a)–(c) depict the variation of: (a) the numbers of nodes and edges with constant average degree $\langle k \rangle = 8$, (b) the edge density for a constant number of nodes $N = 50$, and the stubbornness parameter (c). The quantities $\langle \tilde{\mu}_x \rangle$ and $\langle \tilde{D}_{\max} \rangle$ are computed as averages over 5000 WS networks.

influences $\langle \tilde{\mu}_x \rangle$ and $\langle \tilde{D}_{\max} \rangle$ [244]. For details on the randomization procedure see App. A. In particular, we rewire an increasing number (n_r) of randomly chosen edges and compute both measures over 5000 instances of correspondingly rewired networks. The "original" friendship network is shown at the top right panel of Fig. 4.7. It is characterized by a community structure, resulting from three school classes of a U.S. high school and consists of $N = 70$ nodes and $N_E = 274$ edges. Upon the randomization of edges, the original network is gradually losing its community structure. An exemplary randomization with $n_r = 200$ is shown in the bottom right panel of Fig. 4.7.

Increasing the number of rewired edges reduces $\langle \tilde{D}_{\max} \rangle$, which monotonically decreases. On average the rewiring procedure reduces the MRDs between most stubborn agents by dissolving the community structure. In the case of WS networks, a similar effect was observed, where $\langle \tilde{D}_{\max} \rangle$ decreases monotonically with increasing p_{ws}, which introduces random shortcuts into the ordered structure of a ring lattice. The behavior of $\langle \tilde{\mu}_x \rangle$ is different. After a steep initial increase, $\langle \tilde{\mu}_x \rangle$ reaches a maximum at around $n_r = 40$. Interestingly, a similar maximum arises in the top panel of Fig. 4.6(b), where for $N_E = 150$, two pronounced local maximum values of $\langle \tilde{\mu}_x \rangle$ arise. Such effects may stem from the finite-size of the system, as they are observed for rather small numbers of edges, i.e., the WS network with $N_E = 150$, and the discussed friendship network ($N_E = 270$).

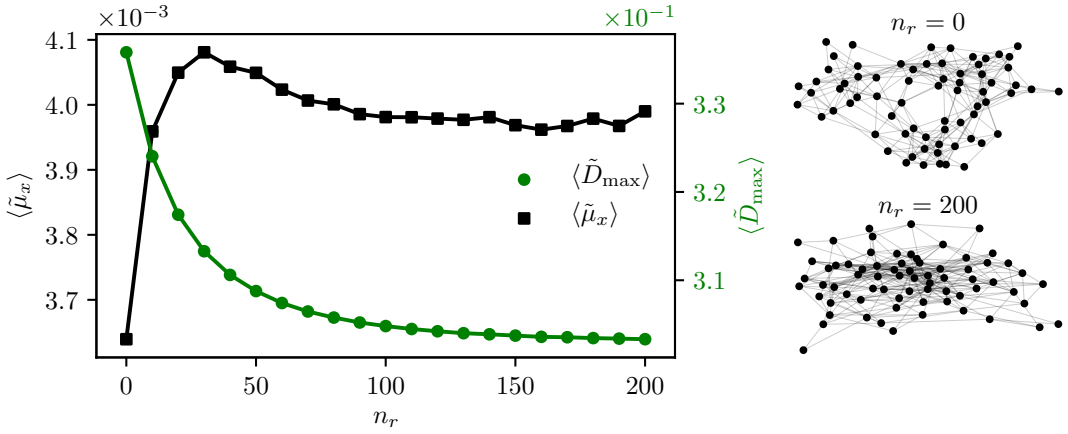

Figure 4.7: Mean absolute opinion $\langle\tilde{\mu}_x\rangle$ and maximum opinion distance $\langle\tilde{D}_{\max}\rangle$ for an empirical friendship network as a function of the number of rewired edges n_r (left panel). The right panel depicts the original network for $n_r = 0$ (top) and after rewiring $n_r = 200$ edges (bottom). For $n_r > 0$ the quantities $\langle\tilde{\mu}_x\rangle$ and $\langle\tilde{D}_{\max}\rangle$ are computed as averages over 5000 networks.

4.6 Chapter summary and discussion

In this chapter, we considered a simplified version of the Taylor model as a linear, and therefore analytically tractable, approach to opinion assimilation processes which are decisively shaped by subsets of stubborn agents. While regular agents aim to continuously minimize the opinion differences to their connected neighbors, stubborn agents are additionally biased towards distinct opinions. We investigated two prototypical situations: (i) the change of a prevailing consensus due to a single stubborn agent, and (ii) the emergence of heterogeneous opinion states for two antagonistically biased agents. Instead of numerically integrating the systems of coupled differential Eqs. (4.2) on a given network topology $\mathbf{W}$, we reformulate the cases of interest, (i) and (ii), in terms of quantities derived from the modified Laplacian $\mathbb{L}^{(\kappa)}$. By that we treated the dynamics of the Taylor model and the topology of the underlying influence network in a unified framework. It allows us to derive compact expressions quantifying relevant features of the emerging opinion dynamics, in terms of MRDs; a novel network metric, which borrows its interpretability from the resistance distance, originally introduced for resistor networks.

While the presented results establish novel links between the opinion dynamics of diffusively coupled agents and the underlying network topology, they are also in line with previous findings. The coherence of consensus change strongly depends on both the network topology and the specific placement of the biased agent. This was demonstrated on the level of single nodes, using $\mathscr{C}$, and on a network level, by the averaged quantity $\tilde{\mathscr{C}}$. The coherence measure decreases for stubborn agents placed on high-degree nodes. Similarly, it has been shown in Ref. [240], using a model for stochastic follower-leader systems, that single leaders with low resistance distances to the remaining agents in the system, i.e., well-connected ones, are followed most closely.

The behavior of $\tilde{\mathscr{C}}$ on WS networks can be contrasted to previous findings on consensus formation on small-world networks, using the majority rule model [78]. It was shown that the time to reach a full consensus in the system decreases strongly and monotonically with the introduction of shortcuts, which have two important effects: (i) geometrically, they reduce the mean shortest distances in the network, and (ii) dynamically, they dissolve local minority groups. While the linear Taylor model strongly differs from the majority rule model, the utilized Laplacian framework, allows to combine those distinct viewpoints. Geometrical and dynamical aspects are simultaneously captured by MRDs in the system. Overall, those are reduced upon the introduction of random links for $p_{\mathrm{ws}} > 0$.

Our results obtained for SBM networks can be compared to previous findings on the effects of community structure on opinion formation processes and cultural dynamics [79, 80]. In Ref. [79], using the naming game, it was found numerically, that committed agents are increasingly hampered to change a status-quo social consensus in the case of a pronounced community structure. Using a different model for information accumulation [80], it was shown, in line with Ref. [79], that a decrease in inter-community connectivity hampers the establishment of a global consensus. Similarly, in the Taylor model a consensus change happens less coherently for a pronounced community structures, as demonstrated for SBM networks.

While it was not possible to express $\mathscr{C}$ (Eq. (4.16)) in terms of MRDs, it could be achieved for the measures D_{max} and σ_x^2. The obtained results for D_{max} and σ_x^2 yield a straightforward interpretation on the network level. Evidently, Eq. (4.21) shows that the largest observed opinion difference in the system is proportional to the MRD between the two opposed stubborn agents. This result is directly reflected Fig. 4.5, where both measures decrease (increase) for increasing values of p_{ws} (p_{intra}). While the introduction of shortcuts reduces the MRD between two random agents, it is strongly increased for a pronounced community structure. Note that, the strong increase of $\tilde{D}_{\mathrm{max}}$, and ultimately its divergence, arises due to cases in which the opposed stubborn agents reside in different communities. In this limit the MRD between two disconnected communities is infinite. Arguments along these lines explain the behavior of σ_x^2. To see this we may consider, for simplicity, that both biased agents a and b, reside on nodes with a similar centrality. In this case we have $\left[C_1^{-1}(b) - C_1^{-1}(a)\right]^2 \simeq 0$, yielding $\sigma_x^2 \sim \Omega_{ab}^{(\kappa,2)}$ for a constant system size N. Thus, both measures of opinion heterogeneity (D_{max}, σ_x^2) are proportional to the MRD (of first- and second-order) between both stubborn agents. In other words, the higher the mutual influence between the opposed agents, i.e. the lower the corresponding MRD, the less polarizing is their overall effect. This may not only be concluded from Fig. 4.5, but also from Fig. 4.6(b), where $\tilde{D}_{\mathrm{max}}$ is increased for a decreasing number of edges in the network. Interestingly, this finding is in line with results obtained for the voter model, where the influence of the zealots' positions in the network on the final opinion distribution was investigated [231]. Specifically, it was found that if two opposing zealots are adjacent, their polarizing effect is substantially decreased. Forming a "dipole" both zealots screen each other and the system shows a preference for consensus on one of the sides. Instead, if both zealots are far apart, a strong polarization is induced, and the population can be divided in two

distinct opinion groups. In the Taylor model, the heterogeneity of opinion states decreases with increasing proximity of stubborn agents, as the MRD between them decreases.

Of course our work comes with limitations. An important shortcoming is that the numerical evaluation of some derived expressions is computationally expensive. In particular, this concerns the networked averaged quantities defined in Eq. (4.17) and Eqs. (4.24)(a)-(c). In App. A, we plot the times needed to compute $\langle\tilde{\mathscr{C}}\rangle$ and $\langle\tilde{D}_{\max}\rangle$ on networks of different sizes, which scale as N^3 and N^5, respectively. Thus our approach is most useful for cases, which do not require such extensive averaging procedures as the positions of stubborn agents are known, or their unknown positions are restricted to a small subset of nodes. Finally, our theoretical result on the opinion association problem, which only requires a single matrix inversion, is also well suited for the application to large systems.

5 Second-order Consensus Dynamics under Time-Periodic Coupling

This chapter is mainly based on the publication

Fabian Baumann, Igor M. Sokolov and Melvyn Tyloo.
Periodic Coupling inhibits Second-order Consensus on Networks.
Phys. Rev. E 102.5 (2020): 052313.

We thank APS for the kind permission to reuse the contained figures for this dissertation.

Consensus models based on a diffusive coupling scheme between the dynamical states of networked agents do not only arise in the context of opinion dynamics. Instead, models with a similar structure have also been used to describe the dynamics of various other systems, where single units aim to reach a global consensus in a decentralized manner. Notable examples in technical applications range from algorithms for multi-vehicle formation control [207, 208, 209, 212, 213], to distributed task assignments protocols [245, 246].

In this chapter, we focus entirely on linear second-order consensus models [212, 215, 216, 216, 217, 218, 219, 220, 221, 222, 223, 224]. In contrast to first-order systems, the Newtonian dynamics of individual agents is accounted for by a second-order time derivative. It it thus possible to model situations in which not only the states of agents (e.g. their positions x_i) form a consensus, but also their derivatives (e.g. their velocities $\dot{x}_i$) converge to a constant non-vanishing value [215]. For second-order models we uncover, and investigate in detail, a mechanism which is induced by small time-periodic coupling modulations and inhibits the formation of collective consensus states. A spectral decomposition of the network Laplacian reveals that, this inhibition is triggered by a parametric resonance phenomenon for certain intermediate coupling frequencies. This is in stark contrast to the expected behavior for very short and long coupling time scales, where indeed consensus emerges [203]. The theory precisely predicts those resonance frequencies and links them to the Laplacian spectrum of the underlying static backbone network. Additionally, the excitation of the agents' amplitudes

is quantified, which extends the concepts of parametric resonance to the domain of complex networks under time-periodic coupling modulations.

The chapter is structured as follows. In Sec. 5.1, we present a second-order consensus model with time-periodic coupling modulations, and discuss the theoretical approach to explain the inhibition of consensus due to parametric resonance on a network level. In a second step the theory is numerically evaluated on different networks (Sec. 5.2). The results are discussed in Sec. 5.3.

5.1 Model and theory

The vast majority of consensus research has focused on static networks [212, 215, 216, 216, 217, 220, 221, 222, 223, 224]. Exceptions include investigations of multi-agent systems on switching topologies [218, 219], where the interaction network changes in discrete steps. It was found that in contrast to first-order systems consensus is not necessarily achieved if infinitely many unions of consecutive network snapshots have a directed spanning tree [219]. In the following, we will not consider second-order models with switching topologies. Instead, the focus is on time-dependent coupling strengths. In particular, while the system's backbone network (**A**) is assumed to be static the coupling weights (**W**) are periodically modulated. Such time-periodic couplings have previously been investigated with respect to the synchronization behavior of non-linear oscillators [247, 248], on which it had a significant impact. As we will show in the following, the formation of second-order consensus is also strongly influenced by similar coupling modulations.

We consider a system of N agents, where the coupling strength between two agents i and j is given by the ijth element of the matrix $\mathbf{W} = f(t)\mathbf{A}$. It is composed of two parts: (i) the *time-independent* and undirected backbone network, encoded by the adjacency matrix **A**, and (ii) the time-periodic modulation function [247, 248]

$$f(t) = 1 + h\sin(\omega t)\,. \tag{5.1}$$

The amplitude and frequency of the modulation are denoted as h and ω, respectively, where it is assumed that $h \in [0,1]$ to exclude repulsive coupling. The static backbone **A** is connected to the coupling matrix via its temporal average, i.e., we have $\mathbf{A} = \lim_{T\to\infty} T^{-1}\int_0^T \mathbf{W}(t)\mathrm{d}t$. The time-dependent state of each agent i (associated to node i in the network) is denoted as $x_i(t)$, where we will omit the time dependence for brevity. The system's dynamics is governed by an established model for second-order consensus on networks [215, 216, 217, 220, 221, 222, 223, 249], which is extended by the sinusoidal coupling modulation defined in Eq. (5.1). In its most

general form, including dissipation [249], the model reads

$$\begin{aligned}\ddot{x}_i + d(t)\,\dot{x}_i = &-\gamma \sum_{j=1}^{N} W_{ij}(t)\,(x_i - x_j) \\ &-\mu \sum_{j=1}^{N} W_{ij}(t)\,(\dot{x}_i - \dot{x}_j)\,,\end{aligned} \tag{5.2}$$

where $W_{ij}(t)$ is ijth element of $\mathbf{W}$ at time t. The coefficients $\gamma > 0$ and $\mu > 0$ control the diffusive coupling strengths between the agents' positions and velocities, respectively, and $d(t)$ is the damping coefficient. In the following, we will consider two different cases of the general model, defined in Eq. (5.2).

On the one hand, we study a system of diffusively coupled damped oscillators, referred to as damped oscillator (DO) model. While agents are assumed to dissipate energy ($d > 0$), they are not coupled via their velocities ($\mu = 0$) and Eq. (5.2) reduces to

$$\ddot{x}_i + d(t)\,\dot{x}_i = -\gamma \sum_{j=1}^{N} W_{ij}(t)\,(x_i - x_j)\,. \tag{5.3}$$

In the DO model, damped harmonic oscillators are coupled via springs with time-varying spring constants $\gamma W_{ij}(t)$. Previously, such networked mass-spring systems have been investigated with time-independent couplings [160]. More importantly, however, Eq. (5.3) corresponds to the linearized dynamics of various nonlinearly coupled systems, with a very wide range of applications, for example, the second-order Kuramoto model, which is a paradigmatic model to investigate the synchronization properties of power grids [158, 250]. In the DO model, consensus is established as a combination of dissipation ($d > 0$) and diffusive position coupling ($\gamma > 0$). Accordingly, if consensus is established all agents will finally come to rest at the same position, i.e., $|x_i - x_j| = 0, \forall\, i, j$ and $\dot{x}_i = 0, \forall\, i$.

On the other hand, we consider the classical second-order consensus (SOC) model, where agents do not dissipate energy ($d = 0$), but align their velocities ($\mu > 0$). In this case, Eq. (5.2) becomes

$$\begin{aligned}\ddot{x}_i = &-\gamma \sum_{j=1}^{N} W_{ij}(t)\,(x_i - x_j) \\ &-\mu \sum_{j=1}^{N} W_{ij}(t)\,(\dot{x}_i - \dot{x}_j)\,,\end{aligned} \tag{5.4}$$

where consensus is defined as a state with $|x_i - x_j| = |\dot{x}_i - \dot{x}_j| = 0, \forall\, i, j$. Note that, in contrast to the DO model, this type of consensus corresponds to a state with equal but (potentially) finite velocities, as discussed in Sec. 3.3.

For a time-independent coupling it can easily be shown that both model versions reach a consensus for $t \to \infty$ [216]. The same holds true for the two limiting cases of $\omega \to 0$ and $\omega \to \infty$

in Eq. (5.1), i.e., for a vanishing and an infinitely large modulation frequency, respectively. While the limit $\omega \to 0$ trivially corresponds to a static coupling matrix, in the second case, the dynamics of the system is effectively governed by the time averaged coupling $\langle \mathbf{W}(t) \rangle$, and thus, also leads to consensus, which can be tested numerically for very high values of ω. Strikingly, for specific intermediate modulation frequencies ω^* the formation of global consensus states is inhibited, and instead agents' amplitudes increase exponentially. We will refer to those intermediate values of ω as resonance frequencies ω^*.

In Fig. 5.1, we show the two models (DO and SOC) in response to both resonant and off-resonant coupling modulation. For $\omega \neq \omega^*$, each model approaches consensus, as shown in the left panel of Fig. 5.1. While velocities vanish for the consensus of the DO model (orange lines), the SOC model preserves the collective motion and all agents move together as one cluster with equal velocities (purple lines). By contrast, for resonant modulation frequencies $\omega \approx \omega^*$ consensus is inhibited and agents' amplitudes diverge for both the DO and the SOC model. It is the goal of this chapter to establish an analytical link between those resonance frequencies ω^* and properties of the static backbone network $\mathbf{A}$ within a spectral Laplacian approach.

To this end, we first express the general model, Eq. (5.2), in vectorial form. This yields

$$\ddot{\mathbf{x}} + d(t)\dot{\mathbf{x}} = -\mathbb{L}(t)\left(\gamma \mathbf{x} + \mu \dot{\mathbf{x}}\right), \tag{5.5}$$

where $\mathbb{L}(t)$ denotes the time-dependent Laplacian, and $\mathbf{x}$, $\dot{\mathbf{x}}$, and $\ddot{\mathbf{x}}$ are the vectors containing the agents' state variables x_i, and their first and second-order time derivatives, respectively. Importantly, the time-dependent Laplacian $\mathbb{L}(t) = f(t)\mathbb{L}^{(0)}$ factorizes into $f(t)$ and the static Laplacian of the backbone network $\mathbf{A}$

$$\mathbb{L}^{(0)}_{ij} = \begin{cases} -A_{ij}, & i \neq j, \\ \sum_k A_{ik}, & i = j. \end{cases} \tag{5.6}$$

as defined in Ch. 2. Accordingly, the time-dependent eigenvalues of $\mathbb{L}(t)$, $\lambda_\alpha(t)$, factorize as $\lambda_\alpha(t) = f(t)\lambda^{(0)}_\alpha$, where $\lambda^{(0)}_\alpha$ denotes the αth eigenvalue of the static Laplacian $\mathbb{L}^{(0)}$. As in the previous Ch. 4, we spectrally decompose Eq. (5.5), using the ansatz $\mathbf{x}(t) = \sum_\alpha c_\alpha(t)\,\mathbf{u}_\alpha$. This yields

$$\ddot{c}_\alpha + k(t)\dot{c}_\alpha + \gamma\lambda_\alpha(t)c_\alpha = 0, \quad \text{for} \quad \alpha = 1, \ldots, N, \tag{5.7}$$

where we have defined $k(t) = d(t) + \mu\lambda_\alpha(t)$. The αth expansion coefficient of $\mathbf{x}(t)$ on the set of eigenvectors $\{\mathbf{u}_\alpha\}$ of $\mathbb{L}^{(0)}$ is denoted as c_α.

In this spectral formulation the effects of $\mathbf{W}(t)$ on the dynamics of the system are particularly evident: the time-periodic coupling corresponds to parameter modulations of damped oscillator equations, which govern the dynamics of network modes α (Eq. (5.7)). Hence, the system is decomposed into N modes, where each mode α corresponds to a one-dimensional parametric

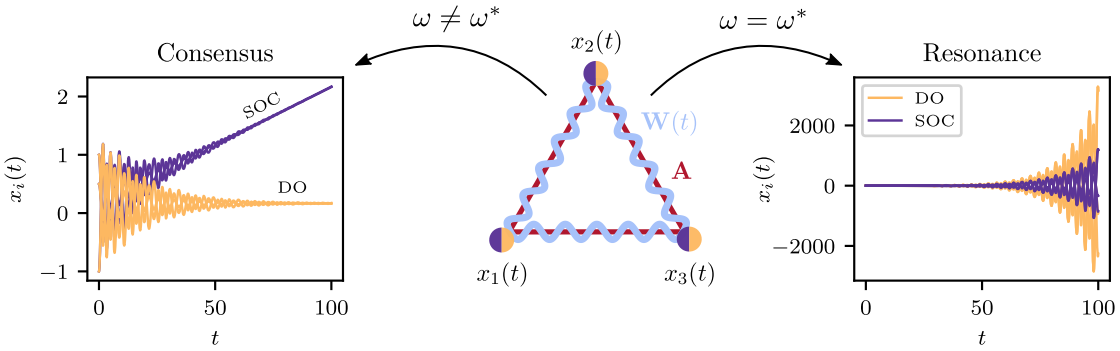

Figure 5.1: Time evolution of the DO (orange) and SOC (purple) model for different modulation frequencies ω. In the left panel, the dynamics for sufficiently off-resonant modulation ($\omega \neq \omega^*$) is shown. Both models approach consensus, with vanishing (DO model) and generally non-vanishing final velocities (SOC model). In contrast, for resonant modulations ($\omega \approx \omega^*$) the formation of consensus is inhibited and leads to an excitation of the system, characterized by exponentially growing amplitudes, as depicted in the right panel. In the center, we schematically show the setup of the time-dependent coupling for a fully connected network of three nodes: while the static backbone network $\mathbf{A}$ (red triangle) does not change over time, the coupling strengths $\mathbf{W}(t)$ are periodically modulated, according to Eq. (5.1) (wavy blue line).

oscillator, which may be excited individually via parametric excitation [251]. Such parametric oscillators are not excited by an explicit external force. Instead, excitation results from the aforementioned periodic variations of the system's parameters. Thus for resonant modulation (for which parametric resonance is obtained) the system's amplitudes increase exponentially [252]. Furthermore note that the expression for $k(t)$ reveals the analogy between dissipation ($d > 0$) and velocity alignment ($\mu > 0$): both mechanisms may contribute to the damping of the αth mode (see Eq. (5.7)), and are therefore expected to contribute to the emergence of consensus in similar ways.

As for $\mathbf{W}(t) > 0$ ($h \in [0,1]$) the dynamics is purely assimilative and the agents' states generally aim to converge it may come as a surprise that consensus is inhibited for certain modulation frequencies. However, periodic modulations of the coupling strength translate into a collective resonance phenomenon induced by parametric excitation, which leads to an exponential growth of the oscillation amplitudes. To shed further light onto this phenomenon, we aim to establish an analytical link between the spectrum of the static Laplacian, $\lambda_\alpha^{(0)}$, and the expected resonance frequencies. First, we simplify Eq. (5.7) using the following variable transformations

$$c_\alpha(t) = e^{-K(t)} q_\alpha(t)\,, \tag{5.8}$$

$$K(t) = \frac{1}{2}\int_0^t k(t')\mathrm{d}t'\,. \tag{5.9}$$

This yields

$$\ddot{q}_\alpha + \Omega_\alpha^2(t) q_\alpha = 0\,, \tag{5.10}$$

the Hill equation [253], where $\Omega_\alpha(t)$ denotes the time-dependent frequency, defined as

$$\Omega_\alpha^2(t) = \gamma\lambda_\alpha(t) - k^2(t)/4 - \dot{k}(t)/2\,. \tag{5.11}$$

Importantly, the time-dependence of $\Omega_\alpha(t)$ either results from the damping coefficient $d(t)$, the eigenvalue $\lambda_\alpha(t)$, or both. Accordingly, Eq. (5.11) takes a different form for each of the two model versions (DO and SOC), which we will therefore consider separately.

Damped oscillator model

Assuming time-independent damping [$d(t) = d$] in the DO model yields $k = d$ and $\dot{k} = 0$. Thus, Eq. (5.11) becomes

$$\Omega_\alpha^2(t) = \gamma\lambda_\alpha^{(0)}\,[1 + h\sin(\omega t)] - d^2/4. \tag{5.12}$$

Plugging Eq. (5.12) in Eq. (5.10) and defining $\omega_\alpha^2 = \gamma\lambda_\alpha^{(0)} - d^2/4$ yields

$$\ddot{q}_\alpha + \omega_\alpha^2\left[1 + \frac{\gamma\lambda_\alpha^{(0)}h}{\omega_\alpha^2}\sin(\omega t)\right]q_\alpha = 0\,, \tag{5.13}$$

which suggests that parametric excitation of the αth mode is expected for

$$\omega_\alpha^* = 2\omega_\alpha = 2\sqrt{\gamma\lambda_\alpha^{(0)} - \left(\frac{d}{2}\right)^2}\,. \tag{5.14}$$

This corresponds to twice the natural frequency of the system ω_α, as it is generally the case for one-dimensional parametric oscillators [252]. Importantly, however, not all eigenvalues $\lambda_\alpha^{(0)}$ will give rise to a resonance frequency; only those eigenvalues fulfilling the following conditions:

$$4\gamma\lambda_\alpha^{(0)} > d^2\,, \tag{5.15}$$

$$\left|\frac{\gamma\lambda_\alpha^{(0)}h}{4\omega_\alpha}\right| > d\,. \tag{5.16}$$

In particular, while condition (5.15) ensures that ω_α^* is purely real (see Eq. (5.14)), Eq. (5.16) implies the exponential growth of the induced oscillations, as we will discuss below.

Assuming small oscillation amplitudes, i.e, $\gamma\lambda_\alpha^{(0)}h/\omega_\alpha^2 \ll 1$, the solution of Eq. (5.10) will take the form of the ansatz

$$q_\alpha(t) = a_\alpha(t)\cos[(\omega_\alpha + \varepsilon)t] + b_\alpha(t)\sin[(\omega_\alpha + \varepsilon)t]\,, \tag{5.17}$$

where ε is the detuning from ω_α^* and terms containing multiples of $(2\omega_\alpha + \varepsilon)$ were neglected [252]. Inserting Eq. (5.17) into Eq. (5.13), for $\omega_\alpha^* = 2\omega_\alpha + \varepsilon$, and neglecting terms of order $\mathcal{O}(h^2)$,

we get

$$\begin{aligned} -2\dot{a}_\alpha + a_\alpha \frac{\gamma\lambda_\alpha^{(0)} h}{2\omega_\alpha} - 2\varepsilon b_\alpha &= 0\,, \\ 2\dot{b}_\alpha + b_\alpha \frac{\gamma\lambda_\alpha^{(0)} h}{2\omega_\alpha} - 2\varepsilon a_\alpha &= 0\,. \end{aligned} \tag{5.18}$$

Searching the solutions to Eqs. (5.18) as $a_\alpha(t)$, $b_\alpha(t) \sim e^{s_\alpha t}$ yields the following expression for the exponent s_α

$$\begin{aligned} s_\alpha^2 &= \left[\frac{\gamma\lambda_\alpha^{(0)} h}{4\omega_\alpha}\right]^2 - \varepsilon^2 \\ &= \left[\frac{\gamma\lambda_\alpha^{(0)} h}{4\sqrt{\gamma\lambda_\alpha^{(0)} - d^2/4}}\right]^2 - \varepsilon^2\,. \end{aligned} \tag{5.19}$$

Resonances for the αth mode are thus expected to be excitable on the following interval of ε,

$$-\frac{\gamma\lambda_\alpha^{(0)} h}{4\sqrt{\gamma\lambda_\alpha^{(0)} - d^2/4}} < \varepsilon < \frac{\gamma\lambda_\alpha^{(0)} h}{4\sqrt{\gamma\lambda_\alpha^{(0)} - d^2/4}}\,, \tag{5.20}$$

centered around the resonance frequency ω_α^*. The width of the resonance interval depends linearly on h, such that the region of excitation is increased for larger modulation amplitudes. For perfectly resonant modulation ($\varepsilon = 0$) the amplitudes $a_\alpha(t)$ and $b_\alpha(t)$ are expected to grow exponentially as $e^{s_\alpha t}$. Together with Eq. (5.8) this yields $c_\alpha(t) \propto e^{(s_\alpha - d)t}$ from which condition (5.16) follows.

Second-order consensus model

Similarly, resonances for the SOC model may be derived. Considering Eq. (5.11), for $k(t) = \mu\lambda_\alpha^{(0)}[1 + h\sin(\omega t)]$, yields

$$\begin{aligned} \Omega_\alpha^2(t) &= \gamma\lambda_\alpha^{(0)} - \left(\frac{\mu\lambda_\alpha^{(0)}}{2}\right)^2 \\ &\quad + \left[\gamma\lambda_\alpha^{(0)} h - 2h\left(\frac{\mu\lambda_\alpha^{(0)}}{2}\right)^2\right]\sin(\omega t) \\ &\quad - \frac{h\omega\mu\lambda_\alpha^{(0)}}{2}\cos(\omega t) - \left(\frac{h\mu\lambda_\alpha^{(0)}}{2}\right)^2 \sin^2(\omega t)\,. \end{aligned} \tag{5.21}$$

For small modulation amplitudes ($h \ll 1$) we may neglect terms of the order $\mathcal{O}(h^2)$ in the expression of $\Omega_\alpha^2(t)$ and therefore obtain

$$\Omega_\alpha^2(t) \simeq \omega_\alpha^2 \left(1 + \frac{B}{\omega_\alpha^2} \sin(\omega t + \varphi)\right) \tag{5.22}$$

with

$$\omega_\alpha^2 = \gamma\lambda_\alpha^{(0)} - \left(\frac{\mu\lambda_\alpha^{(0)}}{2}\right)^2, \tag{5.23}$$

$$B = h\sqrt{\left(\frac{\omega\mu\lambda_\alpha^{(0)}}{2}\right)^2 + \left(\gamma\lambda_\alpha^{(0)} - \frac{\left(\mu\lambda_\alpha^{(0)}\right)^2}{2}\right)^2}, \tag{5.24}$$

and

$$\varphi = \tan^{-1}\left(\frac{\omega\mu\lambda_\alpha^{(0)}}{2\gamma\lambda_\alpha^{(0)} - \left(\mu\lambda_\alpha^{(0)}\right)^2}\right). \tag{5.25}$$

As for the DO model, resonances of the SOC model are found at twice the natural frequencies ω_α. Thus we have

$$\omega_\alpha^* = 2\omega_\alpha = 2\sqrt{\gamma\lambda_\alpha^{(0)} - \left(\frac{\mu\lambda_\alpha^{(0)}}{2}\right)^2}, \tag{5.26}$$

and the growth exponent becomes

$$s_\alpha^2 = \left[\frac{B}{4\omega_\alpha}\right]^2 - \varepsilon^2. \tag{5.27}$$

Accordingly, exponential growth of $c_\alpha(t)$ is expected if $s_\alpha t - K(t) > 0$. The interval, on which a given resonance ω_α^* is excitable, is $\omega \in [\omega_\alpha^* - \varepsilon, \omega_\alpha^* + \varepsilon]$ with $\varepsilon = B/(4\omega_\alpha)$. For the SOC model, the integral of Eq. (5.9) yields

$$K(t) = \frac{\mu\lambda_\alpha^{(0)}}{2}\left(t + \frac{h}{\omega}(1 - \cos(\omega t))\right), \tag{5.28}$$

and reduces to

$$K(t) \simeq \frac{\mu\lambda_\alpha^{(0)}}{2} t \tag{5.29}$$

for small modulation amplitudes ($h \ll 1$). Thus, resonances are expected to appear if the following two inequalities hold:

$$B\omega_\alpha - 2\mu\lambda_\alpha^{(0)} > 0 \tag{5.30}$$

and

$$\gamma - \frac{\mu^2}{4}\lambda_\alpha^{(0)} > 0. \tag{5.31}$$

Note that the velocity alignment mechanism (active for $\mu > 0$) is hampering the emergence of parametric resonance (in the SOC model) in a similar way as the viscous damping does ($d > 0$) in the DO model.

The presented results are rather general and allow to determine the resonance frequencies ω_α^* for arbitrary backbone networks **A**. Below, we will apply the theory on different backbone networks including complete graphs, ring networks, and random networks. While the eigenvalues $\lambda_\alpha^{(0)}$ of $\mathbb{L}^{(0)}$ can be obtained analytically for many regular graphs, the spectra need to be determined numerically in the case of more complex random networks.

5.2 Numerical results

To validate and complement the theoretical analysis of the previous section, we perform numerical simulations. In particular, we obtain resonance spectra $\mathscr{A}(\omega)$, which are generated as follows. We numerically integrate Eq. (5.3) and (5.4) to obtain the time evolution $\mathbf{x}(t)$ of the system for a specific modulation frequency ω. Subsequently, we average the integral $\int_0^T \mathbf{x}^2(t)\mathrm{d}t$ over multiple random initial conditions (IC), yielding the spectra

$$\mathscr{A}(\omega) = \left\langle \int_0^T \mathbf{x}^2(t)\,\mathrm{d}t \right\rangle_{\{\mathrm{IC}\}}, \tag{5.32}$$

which quantify the response of the system to a specific modulation frequency ω, in terms of the agents' integrated squared amplitudes. While consensus gives rise to flat parts in such spectra, resonances manifest as pronounced peaks. The classic case of a static network is recovered for $\omega = 0$, and thus corresponds to $\mathscr{A}(0)$. Without the loss of generality, we initialize the models as resting non-consensus states, i.e. we randomly sample the initial states $x_i(0)$ on the interval $\mathscr{I}_{\mathrm{init}}$ and assume $\dot{x}_i = 0, \forall i$. To facilitate comparability between both model versions we assume ($d = 0.01$, $\gamma = 1$) and ($\mu = 0.01$,$\gamma = 1$) for the DO and SOC model, respectively. Unless indicated differently, we set $\mathscr{I}_{\mathrm{init}} = [-1, 1]$ and $h = 0.2$.

5.2.1 Amplitude growth

To verify the analytical prediction with respect to the exponential amplitude growth for resonant modulation we assume, for simplicity, $d = \mu = 0$. In this case, both considered model versions coincide and the corresponding resonance frequencies are $\omega_\alpha^* = 2\sqrt{\gamma\lambda_\alpha^{(0)}}$ [see Eq. (5.14) and (5.26)]. Accordingly, the exponent s_α becomes

$$s_\alpha = \frac{h\sqrt{\gamma\lambda_\alpha^{(0)}}}{4}. \tag{5.33}$$

In Fig. 5.2, the corresponding parametric excitation is depicted on a complete graph of $N = 100$ nodes. The trajectory of a single node and its local maximum values are depicted in green,

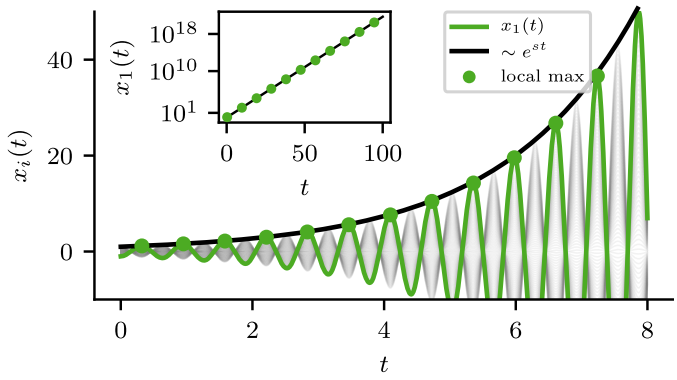

Figure 5.2: Exponential amplitude growth of $x_i(t)$ in response to resonant modulation ($\omega = \omega^*$). The damping and velocity alignment parameters were set to zero, i.e., $d = \mu = 0$ in Eq. (5.5) and we consider a complete graph with $N = 100$ nodes. The growth of the system's amplitudes is captured by the dynamics of the αth expansion coefficient $c_\alpha(t) \sim e^{st}$ (black line). The green line depicts the trajectory of a single agent i together with its local maximum values (dots). The inset shows the amplitude growth for the same agent (green dots) in comparison with the theoretical prediction for a longer simulation time over many orders of magnitude.

while all remaining nodes' trajectories are shown in gray. As predicted from the theory of parametric excitation, the system's amplitudes increase exponentially. Thus, the growth is well captured by $c_\alpha(t) \sim e^{s_\alpha t}$ (black line) not only for short but also long times over many orders of magnitude, as shown in the inset of Fig. 5.2.

5.2.2 Complete graphs

As a first step, we verify the theoretical predictions of resonance frequencies on complete graphs, where each node is connected to all other nodes in the network – a network type, which is often used in studies of coupled oscillators [250]. The Laplacian eigenvalues of a complete graph of N nodes are given as $\lambda_0^{(0)} = 0$ and $\lambda_1^{(0)} = N$, where $\lambda_1^{(0)}$ has the algebraic multiplicity $N-1$ [254]. Consequently, parametric resonance is expected around

$$\omega^* = 2\sqrt{\gamma N - \left(\frac{d}{2}\right)^2} \tag{5.34}$$

for the DO model, and

$$\omega^* = 2\sqrt{\gamma N - \left(\frac{\mu N}{2}\right)^2} \tag{5.35}$$

for the SOC model. Note that, on complete graphs there is (at most) a single resonance frequency for each model, namely the one which corresponds to the eigenvalue $\lambda_1^{(0)} = N$.

Figure 5.3 shows the respective resonance spectra $\mathscr{A}(\omega)$ for a complete graph of $N = 5$ nodes. Here, as well as in all following figures, the orange and purple lines show the spectra of

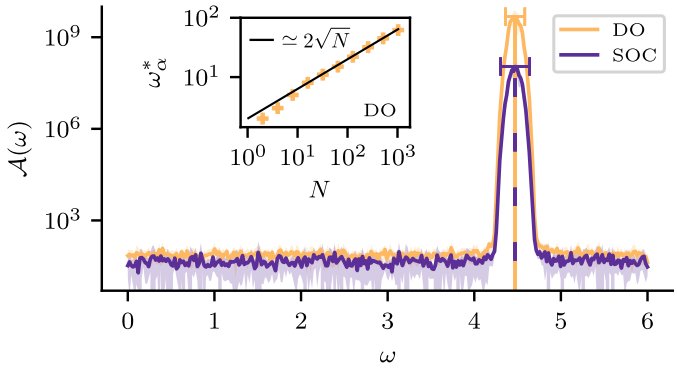

Figure 5.3: Resonance spectra $\mathscr{A}(\omega)$ on a complete graph of $N = 5$ nodes. The vertical orange and dashed purple lines locate the theoretically predicted resonance frequencies for the DO and the SOC model, respectively. The spectrum of the corresponding model is shown in the same color on the interval $\omega \in [0,6]$. The spectra are averaged over 10 runs with the shaded areas indicating the corresponding standard deviations for a simulation time of $T = 100$. The widths of the peaks are well described by the detuning intervals (ε), cf. Eq. (5.20) and Eq. (5.27). The inset shows the predicted resonance frequency $\omega^* \simeq 2\sqrt{N}$ for the DO model compared to simulations for systems sizes up to $N = 10^3$.

the DO and the SOC model, respectively. Vertical lines depict the theoretically predicted resonance values, and are colored according to the corresponding model. For both models the resonance peaks are matched well by the predicted values of ω^*, and the peak widths are in good agreement with the theoretical detuning ranges (ε). As Eq. (5.34) and (5.35) and Fig. 5.3 suggest, both models behave similarly for small values of d and μ and small systems sizes. The thermodynamic limit of $N \to \infty$, however, yields a different behavior, where no resonance may emerge for the SOC model if the values of γ and μ are fixed [see (5.35)]. By contrast, the DO model will still give rise to resonances in the case of $N \to \infty$, and the corresponding frequencies are well described by $\omega^* \simeq 2\sqrt{N}$ (with $\gamma = 1$), as supported by numerical simulations for network sizes up to $N = 1000$ (see inset of Fig. 5.3). We conclude that for growing system sizes the consensus of the DO model becomes increasingly robust with respect to low frequency modulation, while high values of $\omega \sim \omega^*$ may still induce resonant behavior and therefore inhibit consensus formation.

5.2.3 Ring networks

In contrast to the previous case of complete graphs, the Laplacian spectrum of ring networks may contain more than a single non-vanishing eigenvalue. This property potentially leads to multiple resonances at different modulation frequencies. Specifically, the Laplacian eigenvalues of ring networks distribute on the interval $\lambda_\alpha^{(0)} \in [0,4]$. The αth eigenvalue is given by

$$\lambda_\alpha^{(0)} = 2 - 2\cos(k_\alpha), \tag{5.36}$$

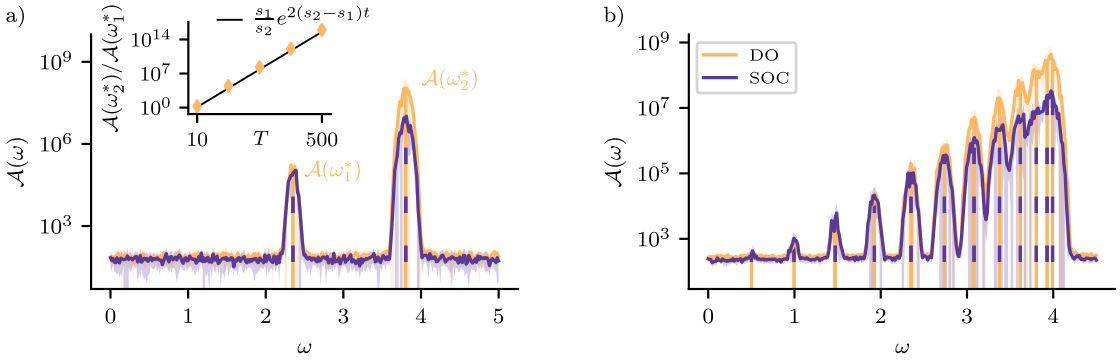

Figure 5.4: Resonance spectra $\mathscr{A}(\omega)$ on ring networks. The vertical orange and dashed purple lines locate the theoretically predicted resonance frequencies for both models (DO and SOC) on ring networks of $N = 5$ (a) and $N = 25$ (b) nodes, respectively. The spectra are averaged over 10 runs with the shaded areas indicating the corresponding standard deviations, with a simulation time of $T = 100$. The inset in panel (a) shows the ratio $\mathscr{A}(\omega_2^*)/\mathscr{A}(\omega_1^*)$ for the DO model as a function of the simulation time T (orange rhombuses), which is well described by the theory, i.e., $e^{2(s_2-s_1)t}$ (black line), where s_1 and s_2 are given by Eq. (5.19).

with $k_\alpha = 2\pi(\alpha-1)/N$ [254]. From this follow the resonance frequencies as

$$\omega_\alpha^* = 2\sqrt{2-2\cos(k_\alpha)-\left(\frac{d}{2}\right)^2} \tag{5.37}$$

for the DO model [see Eq. (5.14)]. Likewise, the resonances of the SOC model are given by Eq. (5.26) as

$$\omega_\alpha^* = 2\sqrt{2-\mu^2+(2\mu-2)\cos(k_\alpha)-\mu^2\cos^2(k_\alpha)}\,. \tag{5.38}$$

In Fig. 5.4(a), we show the spectra for the case of a ring network of $N = 5$ nodes. Note that, for $d = \mu = 0.01$, both distinct eigenvalues $\lambda_\alpha^{(0)} > 0$ give rise to a resonance, as the corresponding resonance conditions are satisfied. The two peaks in the spectra are matched by their theoretical predictions for both models. For a simulation time of $T = 100$ the peak heights are separated by several orders of magnitude, where the separation between $\mathscr{A}(\omega_1^*)$ and $\mathscr{A}(\omega_2^*)$ is larger for the DO model. In the inset of Fig. 5.4(a), we depict the ratio $\mathscr{A}(\omega_1^*)/\mathscr{A}(\omega_2^*)$ for the DO model as a function of the simulation time T. It is captured well by the theory over many orders of magnitude. For large T, the ratio can be approximated as $\int c_2^2 \mathrm{d}t / \int c_1^2 \mathrm{d}t \sim (s_1/s_2)\, e^{2(s_2-s_1)T}$, which is obtained from Eq. (5.19) and shown as black line in the inset of Fig. 5.4(a).

In Fig. 5.4(b), the spectra for a ring network of $N = 25$ nodes are shown, where each model version gives rise to twelve resonances. The results reveal an important consequence of the bounded Laplacian spectrum of ring networks. Instead of ever increasing values of ω^* for growing system sizes, as for the DO model on complete graphs, we observe that for $N \gg \infty$, the bounded intervals of resonance frequencies are increasingly filled. As a result,

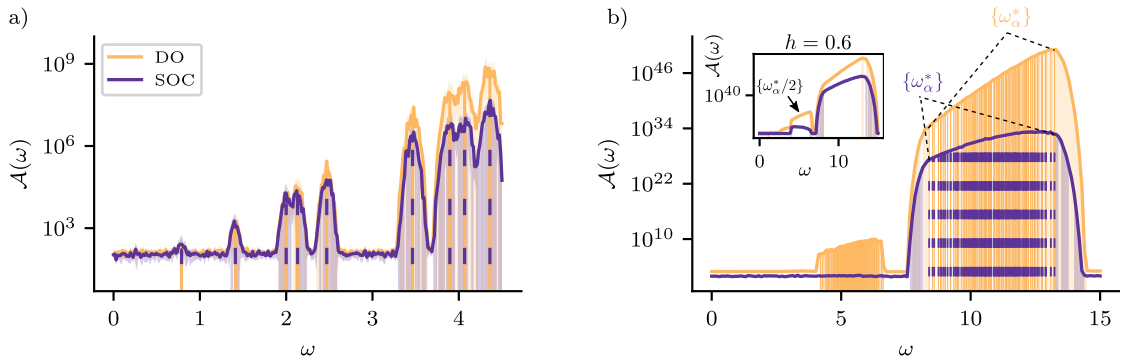

Figure 5.5: Resonance spectra $\mathscr{A}(\omega)$ for Erdős-Rényi networks. The vertical orange and dashed purple lines locate the theoretically predicted resonance frequencies for both models (DO and SOC) on ER networks of $N = 25$ (a) and $N = 100$ (b) nodes, respectively. The spectra are averaged over 10 runs with the shaded areas indicating the corresponding standard deviations for a simulation time of $T = 100$. In panel (b) the modulation amplitudes are set to $h = 0.4$ and $h = 0.6$ in the main plot and the inset, respectively.

also intermediate frequencies may be excited if two resonance frequencies are within their respective ε-ranges. The effect is especially strong for $4 < \omega \leq 4.5$, where the increased values of $\mathscr{A}(\omega)$ between two predicted peaks at ω_α^* and $\omega_{\alpha+1}^*$ indicate a partially off-resonant excitation of both models.

5.2.4 Random networks

The previous two cases of complete and rings networks verified the theory on highly ordered network structures. The question is if resonance frequencies can be reliably predicted also on random networks obeying disorder. In this section, the focus is therefore on Erdős-Rényi (ER) networks, where the Laplacian spectra need to be computed numerically.

The results are shown in panels (a) and (b) of Fig. 5.5, where we depict $\mathscr{A}(\omega)$ for two ER networks of $N = 10$ (a) and $N = 100$ (b) nodes, respectively. Especially, the case of the small ER network reveals the random network structure, which is reflected in the irregular positions of the resonance peaks resulting from the random spectrum of $\mathbb{L}^{(0)}$. Despite the inherent disorder of the ER network, the resonance frequencies ω_α^* are successfully predicted for both models.

The spectrum of the larger ER network ($N = 100$) is shown in panel (b) of Fig. 5.5. Here, the density of resonance frequencies is strongly increased and the detuning ranges ε of single resonances overlap, which leads to a quasi-continuous excitation of the system on the interval of approximately $\omega \in [8, 13]$. The spectrum of the DO model shows a further feature, known for general parametric oscillators, namely the excitation of the system at half the resonance frequency [252]. The inset of Fig. 5.5 demonstrates that this effect becomes increasingly

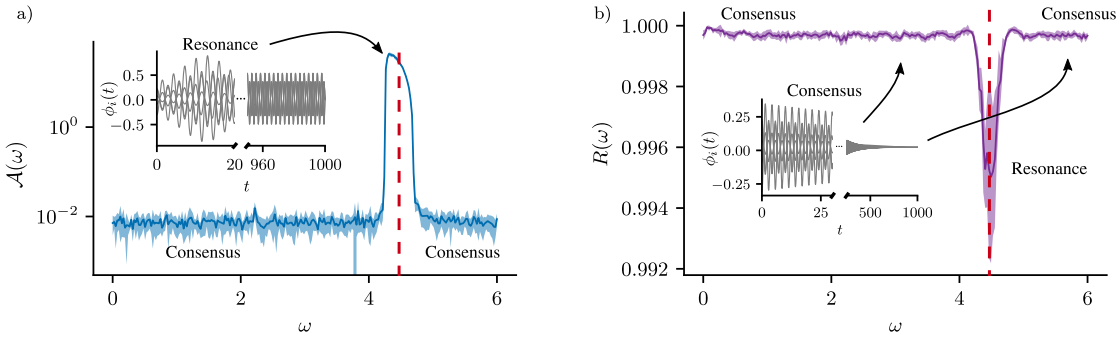

Figure 5.6: Resonance spectrum $\mathscr{A}(\omega)$ and order parameter for the second-order Kuramoto model as a function of the modulation frequency ω. Panel (a) depicts the time-evolution of the second-order Kuramoto model for resonant coupling modulation ($\omega^* \simeq 4.3$) (inset) and the spectrum $\mathscr{A}(\omega)$ on the interval $\omega \in [0,6]$ (main plot). The red dashed line locates the predicted resonance frequency ω^* for a complete graph of $N = 5$ nodes close to the consensus fixed point. The shaded blue area shows the standard deviation of $\mathscr{A}(\omega)$ for 10 random initial conditions. Panel (b) depicts the Kuramoto order parameter R plotted against the modulation frequency ω (main plot) and the time-evolution of the second-order Kuramoto model for off-resonant coupling modulation, for which consensus emerges.

pronounced for larger modulation amplitudes ($h = 0.6$). For this case, we furthermore find that half-frequency excitations are generally not limited to the DO model, but also appear for the SOC model.

5.2.5 Non-linear oscillators

We derived the theory of parametric resonance induced by time-periodic couplings for linear second-order consensus models of diffusively coupled units [see Eq. (5.5)]. However, Eq. (5.5) may also be interpreted as the linearization of certain models with non-linear coupling. Our formalism may therefore also be applied to investigate the onset of parametric resonance in non-linear systems close to a stable fixed point.

Traditionally, one of the most studied models of non-linear second-order oscillators is the second-order Kuramoto model, due to its applications in the study of power grids [158, 250]. Its equations are recovered from Eq. (5.5) by exchanging the diffusive coupling by a sinusoidal one and switching to the common notation for phase oscillator models $x_i \to \phi_i$:

$$\ddot{\phi}_i + d\,\dot{\phi}_i = \omega_i - \sum_{j=1}^{N} W_{ij}(t)\sin(\phi_i - \phi_j)\,, \tag{5.39}$$

where we have set $\mu = 0$ and introduced the natural frequency ω_i of oscillator i. Without loss of generality, we assume $\sum_i \omega_i = 0$, which corresponds to the transformation into a rotating reference frame, yielding a synchronous state with $\dot{\phi}_i = 0\ \forall i$. We denote the ith coordinate of

a stable fixed point $\boldsymbol{\phi}_0$ of Eq. (5.39) on a given network $\mathbf{A}$ as $\phi_{i,0}$, and get

$$\omega_i = \sum_{j=1}^{N} W_{ij}(t)\sin(\phi_{i,0} - \phi_{j,0})\,. \tag{5.40}$$

For $h \ll 1$, the sinusoidal coupling term may be linearized around $\boldsymbol{\phi}_0$. Keeping only terms up until the first order, we get

$$\sin(\phi_{i,0} - \phi_{j,0}) \simeq \cos(\phi_{i,0} - \phi_{j,0})(\phi_i - \phi_{i,0}) \tag{5.41}$$

$$- \cos(\phi_{i,0} - \phi_{j,0})(\phi_j - \phi_{j,0}). \tag{5.42}$$

Considering the dynamics of the deviation from the fixed point, $\delta\phi_i(t) = \phi_i(t) - \phi_{i,0}$, we get

$$\delta\ddot{\phi}_i + d\,\delta\dot{\phi}_i = -\sum_{j=1}^{N} W_{ij}(t)\cos(\phi_{i,0} - \phi_{j,0})(\delta\phi_{i,0} - \delta\phi_{j,0})\,. \tag{5.43}$$

In vectorial form, the latter equation can be written as

$$\delta\ddot{\boldsymbol{\phi}} + d\delta\dot{\boldsymbol{\phi}} = -\mathbb{L}_{\mathrm{w}}(t;\phi_0)\,\delta\boldsymbol{\phi}\,, \tag{5.44}$$

with $\mathbb{L}_{\mathrm{w}}(t;\phi_0) = f(t)\mathbb{L}_{\mathrm{w}}^{(0)}$, where $\mathbb{L}_{\mathrm{w}}^{(0)}$ denotes the weighted and static (time-independent) Laplacian defined as

$$[\mathbb{L}_{\mathrm{w}}^{(0)}]_{ij} = \begin{cases} -A_{ij}\cos(\phi_{0,i} - \phi_{0,j})\,, & i \neq j\,, \\ \sum_k A_{ik}\cos(\phi_{0,i} - \phi_{0,k})\,, & i = j\,. \end{cases} \tag{5.45}$$

This weighted Laplacian takes into account the angle difference $(\phi_{i,0} - \phi_{j,0})$ of two connected oscillators at the fixed point $\boldsymbol{\phi}_0$. Thus, we can apply the theory to Eq. (5.44), by exchanging $\{\lambda_\alpha^{(0)}\}$ with the eigenvalue set of the weighted Laplacian $\{\lambda_{\mathrm{w},\alpha}^{(0)}\}$. In the specific case of $\omega_i = 0$, $\forall$ i, the synchronized state satisfies $\phi_{0,i} = \phi_{0,j}$, $\forall\, i, j$, and the weighted Laplacian reduces to the one used in the linear model, i.e. $\mathbb{L}(t,\phi_0) = \mathbb{L}(t)$.

The insets of Fig. 5.6(a) and (b) show the time-evolution of ϕ_i, according to Eq. (5.39), for resonant (a) and off-resonant modulation (b) on a complete graph of $N = 5$ nodes and the identical damping value as used for the DO model ($d = 0.01$). By reducing the initial interval to $\mathscr{I}_{\mathrm{init}} = [-0.01, 0.01]$ we initialize the system close to the fixed point $\phi_{i,0} = 0$. For off-resonant coupling, the system approaches a global consensus, as shown in the inset of panel (b). In the case of resonant modulation, consensus is inhibited and the system gives rise to stable oscillations with fixed amplitudes. This behavior results from the boundedness of the sinusoidal coupling, and is in contrast to the excitation of the linear models, where amplitudes grow exponentially. Nevertheless, the resonance is captured in the spectrum $\mathscr{A}(\omega)$, as a pronounced peak around ω^*, which is predicted within the linear theory (vertical red and dashed lines). Typically, the collective dynamics of Kuramoto models is quantified by the order parameter R, defined as $R = N^{-1}\sum_j e^{i\phi_j}$ [250]. In panel (b) of Fig. 5.6, R is depicted as

a function of the modulation frequency ω. For off-resonant modulation, where the system reaches consensus, we find $R \simeq 1$. Instead, for $\omega = \omega^*$ the order parameter is reduced as oscillators are driven out of phase, due to the parametric excitation of the system.

5.3 Chapter summary and discussion

In this chapter, we investigated the effects of time-periodic couplings on the collective dynamics of networked second-order consensus models. More specifically, we uncovered a parametric resonance phenomenon, which arises for certain intermediate coupling frequencies and inhibits the formation of collective consensus states. Importantly, this is in contrast to the expected emergence of consensus for static couplings ($\omega = 0$) or for strongly off-resonant couplings ($\omega \neq \omega^*$). The corresponding resonance frequencies were predicted reliably, based on the eigenvalues of the network Laplacian of the static backbone network **A**. For rather small system sizes and low frequency modulations, as well as both weak dissipation and velocity alignment the resonances of the DO and the SOC model appear in close vicinity [see Eq. (5.14) and (5.26)] and lead to similarly strong excitations. In the case of larger resonance frequencies, however, excitations are stronger for the DO model, as suggested by numerical simulations. The developed formalism can also be applied to networks of nonlinearly coupled second-order Kuramoto oscillators. In contrast to the investigated linear models, resonant modulation in the second-order Kuramoto model gives rise to constant oscillations with finite and fixed amplitudes. Such resonances were correctly predicted if the system was probed close to a synchronous fixed point.

Previously, similar time-periodic coupling schemes have been considered for two different non-linear systems [247, 248]. In particular, the authors of Ref. [247] investigated the effects of time-periodic couplings on the synchronization behavior of networked chaotic oscillators. Specifically, systems of diffusively coupled Rössler oscillators were investigated. Similar to our approach, a measure was defined to detect the degree of synchronization depending on the modulation frequency. Interestingly, it was found that time-periodic coupling modulations improved the synchronizability of the system over a wide range of modulation frequencies. In some intermediate regions of ω, however, desynchronization windows emerged for modulation frequencies close to the intrinsic frequency of individual Rössler oscillators. In Ref. [248] time-periodic couplings were investigated for systems of globally coupled first-order Kuramoto oscillators. Compared to the constant coupling case it was found that time-periodic couplings generally suppress synchronization. More specifically, the degree of synchronization, measured in terms of the Kuramoto order parameter, gradually decreased with decreasing coupling frequencies and increasing modulation amplitudes. Although this finding is similar to our results on the second-order Kuramoto-model, where synchronization may be inhibited due to periodic couplings, the underlying mechanism is different: it is not due to parametric excitation, which cannot be sustained in a first-order system. Instead, the authors show that, the breakdown of synchronization can be attributed to the asymmetric oscillation of the temporal synchronization parameter.

It can be concluded, that the effects of time-periodic couplings are not uniform across different coupled dynamical systems. Instead, the effects strongly depend on the nodal dynamics and the underlying coupling network. On the one hand, synchronization may be enhanced, thus providing new paths to synchronization optimization, as shown by numerical means in Ref. [247]. On the other hand, periodic couplings may deteriorate synchronization [248], or inhibit the formation of consensus states as we showed analytically for linear second-order oscillators. Our results may have important implications for a wide variety of second-order systems, where time-varying couplings yield a more realistic situation than the classic static network approach.

Part II Modeling Opinion Cleavages

6 Modeling Non-Consensus States

The increasing interaction rates within modern societies, partly induced by the participatory nature of online social media platforms [255], have not only changed the way we interact with peers, but also how information is consumed and disseminated. More specifically, the low cost for user engagement and the networked architecture of modern communication infrastructures have increased human interaction rates [8] and facilitated information exchange across geographical and social barriers [5, 7, 256]. Notably, these developments even undercut Milgram's striking finding of the six degrees of separation in (offline) social networks [165]. Already in 2002, it was found in an empirical study of email communication that social searches in large-scale networks could be performed in a median of down to five steps [5]. Subsequently, a study investigating the entire social network of Facebook showed that our world is even smaller with respect to online friendship relations, uncovering an average distance between Facebook users of 3.74 [7]. Recently, in a blog post authored by Facebook researchers even lower values of separation were measured, with a mean degree of separation of 3.57 of users worldwide [256]. Within the boundaries of the classical models of opinion assimilation, where peers always become more similar upon social interactions, such intensified modes of interaction and social influence would inevitably accelerate the emergence of a consensus, even on controversial issues.

This classical prediction is in clear contrast to several old and more recent empirical observations [3, 121]. Originally posed by R. Abelson in 1964, it has been a long standing question in the social sciences, under which assumptions heterogeneous and bimodal opinion distributions emerge as a result of social influence among peers [3, 121]. Heterogeneous opinion distributions have been found in large population samples on a community level [84], but also in small groups [257], where especially the latter finding excludes perfect social segmentation as the only possible explanation for persisting non-consensus states [121]. The analysis of more recent survey data confirms these insights on public opinions with respect to different issues. In particular, it was found that opinions are often polarized, which refers to situations, in which the opinion distribution is strongly bimodal, i.e., characterized by two well separated peaks around the neutral consensus [117, 258]. Opinion polarization was especially found with respect to controversial issues like abortion or global warming [18, 259], and more

recently on social media platforms with respect to political orientation [260, 261], U.S. and French presidential elections [262], or the vaccination debate [15, 263].

In Part II of this thesis, we depart from models of pure assimilative social influence. The goal is to answer Abelson's puzzle on opinion cleavages without relying on the assumptions of (i) disconnected social networks or (ii) additional external forces, such as stubborn agents, see Ch. 4. In this chapter, we first review two extensions to classical assimilation models. By different cognitively inspired mechanisms those models give rise to persisting heterogeneous opinion states, also in the case of connected networks. Subsequently, we introduce a minimal model of opinion dynamics based on a different set of assumptions: a simple social reinforcement mechanism and social interactions ruled by homophily. The model successfully reproduces various stylized facts present in empirical data on public opinion formation. More specifically, in Ch. 7 the model is contrasted to Twitter data containing information about the opinions of Twitter users towards different topics, their activities, and their social interaction networks. In Ch. 8, we extend the model to multiple dimensions. We discuss under which conditions, on top of opinion bi-polarization, ideological states emerge, similar to those found in empirical data of the American National Election Studies (ANES).

6.1 Similarity bias

Up to this point, in each model it was assumed that there is always social influence between two agents if they are connected in the social network. This general assumption, however, may be complemented by additional ones. Specifically, it may be assumed that social influence depends on agents' similarity [3]. Essentially, the basic assumption of such similarity based mechanisms is that two agents do not influence each other if their opinion difference is too large. In other words, agents only take into account the opinions of similar other agents. Such similarity biases have been implemented in various models of opinion dynamics [3, 38, 264, 265].

In the following, we will motivate the implementation of similarity biases based on the theory of confirmation bias. As discussed in Ch. 1, research on confirmation biases challenges the idea of unconditional social influence between individuals. In particular, it is argued that a piece of information, such as the message of a peer, may be favored, or taken into account at all, if it confirms individuals' pre-existing beliefs. In turn, it is suggested that information which strongly deviates from one's preconceptions is completely disregarded. For consistency with Part I, we will focus on the implementation of a similarity bias on top of classical opinion assimilation models. In contrast to unconditional social influence, these extensions typically involve a non-linear social influence function, which depends on agents' opinions. Previously, such non-linearities were typically implemented as a sharp threshold, where pairwise social influence only takes place if agents' opinions lie sufficiently close to each other, i.e., agents similarity was defined based on their opinion distance [264, 265].

The corresponding threshold of social influence is usually referred to as *confidence bound*.

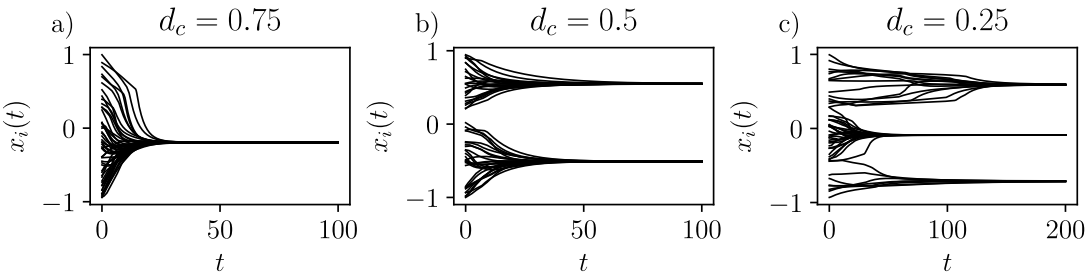

Figure 6.1: Opinion clustering resulting from Eq. (6.1) and Eq. (6.2) on an ER network with $N = 50$ and $p = 0.35$. The initial opinions are sampled randomly and uniformly from the interval $x_i(0) \in [-1, 1]$. Panels (a)-(c) show the situations for different values of the confidence bound: $d_c = 0.75$ (a), $d_c = 0.5$ (b), and $d_c = 0.25$ (c).

Two of the most well-known confidence bound models were introduced about twenty years ago, independently by G. Deffuant [264], and by a collaboration of R. Hegselmann and U. Krause [265]. While Deffuant's model is formulated as a stochastic system based on random pairwise encounters of agents, Hegselmann and Krause proposed a deterministic approach, where all agents update their opinions simultaneously [266]. Following the deterministic approach and assuming an underlying social interaction network $\mathbf{W}$, we formalize the concept of a confidence bound by extending Eq. (3.6). In particular, a non-linear social influence function $f(x_i, x_j)$ is introduced, such that the opinion of agent i evolves according to

$$\dot{x}_i = -\sum_j f(x_i, x_j) W_{ij} \left(x_i - x_j\right), \tag{6.1}$$

where $f(x_i, x_j)$ is defined as a step function, i.e.,

$$f(x_i, x_j) = \begin{cases} 1, & \text{if } |x_i - x_j| \leq d_c \\ 0 & \text{else}, \end{cases} \tag{6.2}$$

and d_c denotes the confidence bound. The non-linearity $f(x_i, x_j)$ profoundly impacts the dynamics of pure opinion assimilation, given by Eq. (3.6). Specifically, if two agents i and j are connected ($W_{ij} = W_{ji} > 1$) the threshold function $f(x_i, x_j)$ gives rise to two different situations. If their opinion distance is within the confidence bound ($|x_i - x_j| \leq d_c$), Eq. (6.2) leads to opinion assimilation, where the rate of opinion convergence is given by $W_{ij} (= W_{ji})$. Otherwise, if the opinion distance exceeds the confidence bound, i.e., ($|x_i - x_j| > d_c$), we have $f(x_i, x_j) = 0$ and *no* social influence is mediated between both agents, although there is a link in the social network between the two: the opinion of agent i does not change due to agent j, and vice versa.

With regard to macroscopic outcomes of confidence bound models, it is instructive to first consider the two limiting cases of vanishing and infinitely large confidence bounds. For $d_c \to 0$

each agent i is maximally intolerant towards other opinions $x_j \neq x_i$, and thus no opinion change may happen, which conserves any initial opinion distribution. By contrast, if $d_c \to \infty$, Eq. (6.2) always yields $f(x_i, x_j) = 1 \, \forall \, i, j$ and all connected agents i and j influence each other. In this case, the confidence bound model reduces to Eq. (3.6), which always leads to consensus for $t \to \infty$. The most interesting dynamical regimes arise for intermediate values of d_c, which we will briefly explore in the following on the basis of Fig. 6.1. Here, the dynamics induced by Eq. (6.2) is shown for three different values of d_c. In each case, the underlying social influence network is an ER network consisting of $N = 50$ nodes, and agents' initial opinions are sampled randomly from the interval $x_i(0) \in [-1, 1]$. For a rather large confidence bound ($d_c = 0.75$) the agents reach a global consensus, shown in panel (a). Decreasing the value d_c reveals the crucial feature of confidence bound models, namely the formation of heterogeneous opinion states which are characterized by *multiple* stable opinion clusters. This situation is depicted for $d_c = 0.5$ [panel (b)] and $d_c = 0.25$ [panel (c)], where two and three persisting opinion clusters emerge, respectively.

While confidence bound models, as defined by Eq. (6.1) and (6.2), may indeed explain the emergence of heterogeneous non-consensus states, they also come with limitations. In the following, we will discuss two of those. First, non-consensus states in confidence bound models are not robust with respect to noise. More specifically, this means that a clustered opinion state, such as the one shown Fig. 6.1(b), dissolves due to either small violations of the confidence bound assumption, where agent i is (with a very low probability) influenced by agents holding opinions outside their confidence bound [267], or random fluctuations of agents' opinions [268]. A second limitation concerns the range of possible final opinions. While confidence bound models may indeed yield more extreme final opinions for some agents, opinions can generally not leave the initial opinion interval [3], which is due to the assimilative opinion dynamics. This is shown in Fig. 6.1, where opinion clusters always form within the initial opinion bounds. This feature is in conflict with empirical findings on group polarization where opinions of like-minded individuals may become more extreme, see Ch. 1. In the following section, we will briefly review another line of research, which extends confidence bound models by a rejection mechanism, due to which agents' opinions can indeed reach values that lie outside of the initial bounds.

6.2 Opinion differentiation

Confidence bound models are based on the assumption that agents stop to influence each other if their opinion difference exceeds a certain threshold value. Accordingly, there is either opinion assimilation or no social influence at all between two agents depending on their relative positions in the underlying opinion space. This view may be complemented by the assumption that two individuals do not always become more similar if they exert social influence on each other. Previous models have therefore proposed differentiation mechanisms, where agents may adjust their opinions to become more dissimilar to their peers. Typically, opinion differentiation was combined with opinion assimilation is such a way that

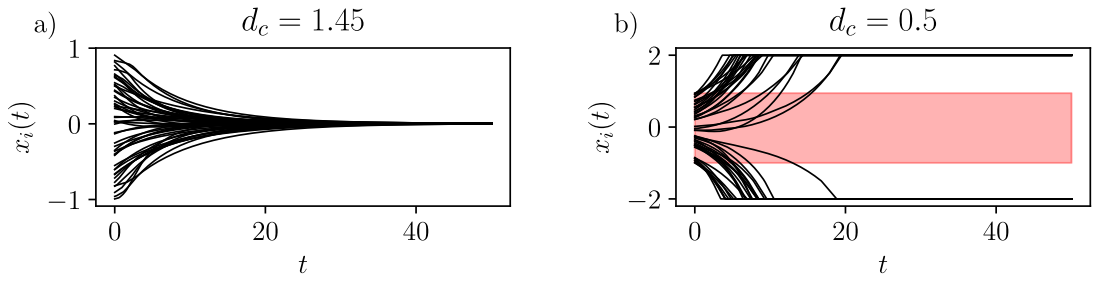

Figure 6.2: Consensus and opinion bi-polarization resulting from Eq. (6.1) and Eq. (6.3) on an ER network with $N = 50$ and $p = 0.35$. The initial opinions are sampled randomly and uniformly from the interval $x_i(0) \in [-1, 1]$. The artificial opinion boundaries were set to $x = -2$ and $x = 2$.

agents tend to become more similar to peers they like or agree with, while they become more dissimilar to peers they disagree with or dislike [3, 109, 269].

Opinion differentiation has been motivated from different angles [3]. First, and most intuitively, it may be assumed that social interactions are (partly) ruled by xenophobia, i.e., the tendency of disliking dissimilar peers [109, 115, 270]. This tendency may induce a repulsion mechanism due to which agents differentiate their opinions from dissimilar peers [271]. Assumptions of opinion differentiation, or opinion repulsion, may also be based on cognitive consistency theories, including Heider's balance theory [35, 36, 109, 270]. As discussed in Ch. 1, balance theory assumes that individuals aim to overcome cognitive "tension states", which are represented by unstable triads involving, themselves, an object and a peer. This balancing procedure may lead to opinion assimilation with similar peers. However, similarly, it may also be argued that individuals aim to differentiate from dissimilar peers in order to overcome cognitive tension.

A prototypical example of opinion differentiation can be formulated on top of the confidence bound model by modifying once more the function $f(x_i, x_j)$. In particular, we define

$$f(x_i, x_j) = \begin{cases} 1, & \text{if } |x_i - x_j| \leq d_c \\ -1 & \text{else .} \end{cases} \tag{6.3}$$

where the first line in Eq. (6.3) recovers opinion assimilation for two connected agents i and j, if their opinions are within the confidence bound d_c. Instead, for opinion differences exceeding the confidence bound, i.e. $|x_i - x_j| > d_c$, the social influence function does not vanish as in Eq. (6.2), but becomes negative. The negative sign of $f(x_i, x_j)$ induces a dynamics in which agents' opinions evolve in opposite directions. In line with previous models [3, 114, 269, 272], we implement an artificial opinion boundary at B_c to avoid the divergence of opinions.

The resulting dynamics is shown in Fig. 6.2, where two different values of d_c lead to drastically

different outcomes. First, in line with regular confidence bound models, the repulsion model recovers the formation of global consensus states for sufficiently large values of d_c, as shown in panel (a). Instead, for smaller values of the confidence bound no global consensus is reached and the repulsion mechanism leads to the formation of two opinion clusters at opposed extreme opinion values, see panel (b). The positions of those opposed opinion clusters are determined by the opinion cut-off, which was set to $B_c = 2$. Note that an important role is played by initially strongly convicted agents. Due to their large opinion difference to most other agents they cause even moderate agents to differentiate from them [3]. Macroscopically, this leads to a situation in which all agents take on extreme opinions outside of the initial bounds, shown as the red shaded area in panel (b).

Despite their theoretical foundations based on concepts such as xenophobia or balance theory, models of repulsive social influence have been criticized due to insufficient empirical evidence [3, 121]. This claim is backed up by a recent empirical study, which found that dissimilarities among individuals did not lead to negative, i.e. repulsive, social influence [273].

6.3 Social reinforcement

The discussed model classes, based on a confidence bound (Sec. 6.1), or repulsive social influence (Sec. 6.2) are essentially (non-linear) extensions to models of opinion assimilation. While they are able to explain the emergence of stable heterogeneous opinion states with a fragmented opinion distribution, they do not take into account phenomena such as group polarization [120, 137] or other social reinforcement mechanisms [54, 80, 274]. Rather recently, the effects of such mechanisms have been explored within different models of opinion dynamics [54, 80, 121, 274].

Figure 6.3: Schematic depiction of opinion assimilation [panel (a)] in contrast to the proposed mechanisms inspired by persuasive-arguments theory [panel (b)]. While two agents decrease their convictions upon an interaction involving two different opinion stances, they increase their convictions in the case of equal opinion stances. This is in contrast to opinion assimilation, where agents' opinions always become more similar.

In the remaining Chs. 7 and 8 of this part, we will introduce a novel and minimal model of social reinforcement. It is based on two different modes of social influence, which are inspired by persuasive-arguments theory (PAT), discussed in Ch. 1. The mechanism will be explained and motivated in detail in Ch. 7. Here we merely want to take the opportunity to distinguish it from opinion assimilation.

In Fig. 6.3, blue and red dots represent the opinions of two interacting agents i and j which

are positive and negative, respectively. In the case of opinion assimilation [panel (a)], agents' opinions always become more similar upon interaction and influence stops as soon as agents opinions are equal. This is different in the case of the mechanism inspired by PAT, see panel (b), where agents always exert social influence on each other if their opinions are non-vanishing. If agents qualitatively disagree on a topic, i.e., if they have different opinion stances ($\mathrm{sgn}(x_i) \neq \mathrm{sgn}(x_j)$), their opinions move towards the neutral state $x = 0$, and thus both reduce their convictions $|x_i|$ and $|x_j|$. By contrast, if their stances are equal ($\mathrm{sgn}(x_i) = \mathrm{sgn}(x_j)$), both agents increase their convictions which models group polarization. We will refer to this latter opinion dynamics towards increasing convictions as radicalization.

7 Radicalization dynamics based on social reinforcement

This chapter is mainly based on the publication

Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov and Michele Starnini.
Modeling Echo Chambers and Polarization Dynamics in Social Networks.
Phys. Rev. Lett. 124.4 (2020): 048301.

We thank APS for the kind permission to reuse the contained figures for this dissertation.

The empirical analyses of large and high-resolved datasets drawn from social networking platforms, such as Facebook or Twitter, provide detailed information on individual users and their mutual interactions. They offer unprecedented opportunities to investigate social dynamics on an individual and a collective level [16, 60, 260, 275, 276]. Empirical research based on such social data have already contributed enormously to our understanding of different aspects of social dynamics in the digital age. Amongst other things, interesting findings include that misinformation spreads faster than factual information [16], social contagion is indeed complex and therefore strongly different from epidemic spreading of most biological pathogens [277], or that information strongly evolves during its dissemination through a social network [278].

Our understanding of opinion dynamics may also benefit substantially from the recent abundance of social data. It is now possible to approximately infer the opinions of a large number of individuals, or users, in an automated way and with respect to different topics, ranging from users' attitudes towards certain commercial products [279] to their political opinions [71, 280, 281]. On top of that – and in contrast to most offline studies on political attitudes or the voting behavior of citizens [90, 282, 283] – data from online social networks often offers the possibility to uncover (at least partially) the relations among users [16, 260, 275], and by that identify sources of social influence. Combined, the information about individuals' opinions and their interactions on social media platforms have revealed old and new characteristics of

collective opinion formation. With respect to politics and other controversial societal or ethical issues, it has been shown that public opinions are often significantly divided, or *polarized*, and therefore strongly deviate from a global consensus [260, 261, 262, 284]. As previously stated, by opinion polarization we refer to situations where the distribution of opinions on a given issue is strongly bimodal, with each side of the issue supported by a large proportion of the population, while moderate or neutral opinions are rare. On a social network level, opinion polarization has recently been found to occur within another prominent signature of various controversial online debates: *echo chambers*, which define situations where one's opinion is correlated to the opinions of one's social contacts, i.e., the opinion of one's neighbors in the social network [85, 86, 87, 88, 89]. Thus, the expressed opinion of a single individual is echoed by their peers. Echo chambers on social media platforms, such as *Twitter* [71, 85, 88, 89] and *Facebook* [15, 285], have raised concerns about their negative implications on the openness of debates within a democratic discourse, and due to their role in the spread of misinformation [15, 16, 87, 286]. In particular, it has been suggested that inside such closed communities of like-minded peers the spreading of information is strongly inclined towards prevailing biases and users are not exposed to diversified pieces of information [15, 85, 286].

The observation of echo chambers as closed groups of like-minded peers may point to a specific mechanism to solve Abelson's puzzle about the origin of bimodal opinion distributions. In this chapter, we introduce a model for opinion dynamics which is strongly different from previous approaches that are based on opinion assimilation extended by a confidence bound or a repulsion mechanism. At its core, the model implements a simple social reinforcement mechanism, which is inspired by persuasive-arguments theory and group polarization (Ch. 1): agents may develop more-extreme opinions upon the interaction with like-minded peers, a process we refer to as *radicalization*. The model couples the opinion evolution to the dynamical social network between agents via homophily, and thus provides a framework where agents' opinions and their social contacts co-evolve. Contrasted with empirical Twitter data the model qualitatively reproduces opinion bi-polarization and echo chambers, as well as the observation that – in polarized debates on social media platforms – more-active users, who are strongly engaged in social interactions, tend to show more-extreme opinions.

The chapter is structured as follows. In Sec. 7.1, the radicalization model is introduced in one dimension. Subsequently, we discuss its dynamical regimes in Sec. 7.2. In Sec. 7.3, the transition from consensus to radicalization is studied within a mean-field approximation. The model is contrasted to empirical Twitter data in Sec. 7.4. Finally, in Sec. 7.5, the chapter is concluded and the results are discussed with respect to previous modeling efforts.

7.1 Radicalization model

Let us consider a system of N agents, where the opinion of agent i is given as the ith element of the time-dependent opinion vector $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^N$. As in previous chapters, we will omit the time-dependence of agents' opinions, $x_i(t)$, for brevity. The sign of x_i, $\mathrm{sgn}(x_i)$,

is defined as the *opinion stance* of agent i. Broadly, it corresponds to the preferred side of an issue, e.g. the preference for one candidate out of two in a presidential election, or the pro/con attitude towards a certain issue or political topic. The absolute value of x_i, $|x_i|$, is the strength of agent i's opinion, or their *conviction*, with respect to the preferred side. Thus, $|x_i|$ quantifies the strength of an agent's stance $\mathrm{sgn}(x_i)$. Accordingly, $x = 0$ with vanishing conviction corresponds to the neutral stance.

We model the dynamics of opinion formation as a collective process which is solely driven by mutual interactions among the agents. Hence, additional factors, such as the influence of mass-media [287, 288, 289], political campaigns [89], opinion leaders [290], or stubborn agents (cf. Ch. 4) are neglected. Specifically, the time-evolution of agents' opinions is described by the following set of N coupled first-order differential equations

$$\dot{x}_i = \underbrace{-x_i}_{\text{finite memory}} \underbrace{+K\sum_{j=1}^{N} A_{ij}(t) f(x_j)}_{\text{social inputs}}, \quad i = 1,\dots,N, \tag{7.1}$$

where $K > 0$ denotes the social interaction strength between two agents, $A_{ij}(t)$ defines the temporal network of interactions, and the function f controls the social influence that agent j exerts on agent i. In contrast to models of opinion assimilation, f only depends on x_j and shall have the following properties: (i) an agent j influences their peers in the direction of their own stance, given by $\mathrm{sgn}(x_j) = \pm 1$, (ii) social influence increases monotonically with agent j's conviction $|x_j|$, and (iii) the social influence of agents with extreme convictions is bounded, as suggested by empirical findings [291]. For concreteness we choose the function

$$f(x) = \tanh(\alpha x), \tag{7.2}$$

whose odd and non-linear shape implements the three properties (i)-(iii). We will explain the interpretation of the parameter α, which tunes the degree of non-linearity in the hyperbolic tangent, a couple of lines below. In the absence of – or for insufficient – social inputs agents' opinions decay towards the neutral state $x = 0$. This assumption is motivated by research on memory processes [292] and was implemented differently in previous models of social influence [80, 121, 270, 293]. Interestingly, similar sigmoid and tunable functions $f(x)$ have previously been used to model non-linear gain functions in models of neural systems, to study chaotic dynamics [294] or the effects of gain on attention and learning [295].

According to Eq. (7.1), the opinion change of agent i is driven by the aggregated social inputs from their neighbors. The social contacts of agent i at time t are determined by the temporal network of social influences $A_{ij}(t)$, with $A_{ij}(t) = 1$ if there is an input from agent j towards agent i, and $A_{ij}(t) = 0$, otherwise. It is assumed that the information flow on social media is generally asymmetric, and the degree of asymmetry depends on the type of social media platform. Therefore, we consider that social influence is established as directed interactions which may be reciprocated with probability r. More specifically, if agent i establishes a

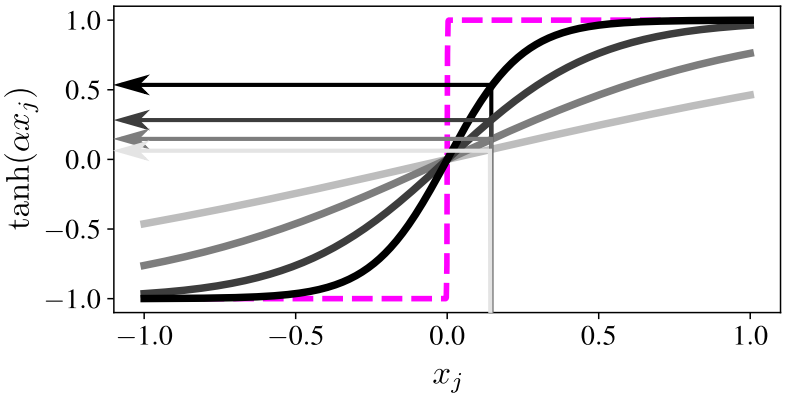

Figure 7.1: The sigmoid social influence function ensures that agents influence their peers in the direction of their own stance, social influence increases monotonically with agents' convictions, and that social influence is capped with respect to extreme opinions. We plotted $\tanh(\alpha x_j)$ for increasing values of α from light gray to black: 0.5, 1, 2, 4. Clearly, the social influence of agents with moderate convictions strongly depends on the controversialness α. The dashed magenta line corresponds to the limit of $\alpha \to \infty$, where $\tanh(\alpha x) \to \mathrm{sgn}(x)$.

connection to agent j, agent j will update her opinion based on x_i, but agent i will do the same only if the interaction is reciprocated.

For a reciprocal interaction between two agents i and j, we distinguish two different situations, which depend on the agents' opinion stances $\mathrm{sgn}(x_i)$ and $\mathrm{sgn}(x_j)$, and have been briefly discussed already in the previous chapter. If both stances are equal $[\mathrm{sgn}(x_i) = \mathrm{sgn}(x_j)]$, the social influence function $\tanh(\alpha x)$ yields an increase of both convictions, referred to as radicalization dynamics. By contrast, for opposite stances $[\mathrm{sgn}(x_i) = -\mathrm{sgn}(x_j)]$, the opinions of agents i and j tend to converge towards the neutral state. Although the mechanism is inspired by persuasive-arguments theory, it is important to remark that we do not model the process of argument exchange explicitly. In contrast to Ref. [121], our model implements PAT in an effective way. While reinforcing social influence between two agents with equal stances mimics the exchange of similar arguments and leads to group polarization, agents' opinions converge towards the neutral state in case of different opinion stances. This latter case models the exchange of opposed arguments. The social influence dynamics induced by the sigmoid non-linearity is similar to previously proposed mechanisms [80, 122, 123]. As in such models – and in contrast to those relying on opinion assimilation – the dynamics of Eq. (7.1) is not shift-invariant. This is due to the distinguished neutral point at $x = 0$. It determines both the opinion stances of two interacting agents and the resulting dynamics between the two: group polarization (for equal stances) or convergence towards the neutral state (for opposed stances).

The amount of social influence that agents exert on their peers is crucially determined by the shape of $\tanh(\alpha x)$, whose degree of non-linearity is tuned by the parameter α. For small α, the effect of moderate opinions x_j on agent i is weak. In contrast, for large α already weakly

convicted agents exert strong social influence on others. These effects are illustrated in Fig. 7.1, where we show the resulting social influence from a moderate agent j for increasing values of α. The darker the color, increasing from light grey to black, the larger is the value of α, as reported in the caption of Fig. 7.1. The dashed magenta line corresponds to the limiting case of $\alpha \to \infty$, where social influence essentially becomes a binary "vote" with $\tanh(\alpha x) \to \mathrm{sgn}(x)$. In this case, convictions do not play a role and agents always exert maximal social influence on others. Loosely speaking, this corresponds to a case of "black-and-white" social influence, where "whoever is not with me is against me" [296]. These considerations lead us to interpret the parameter α as the *controversialness* of the topic discussed. It has been suggested that controversialness can be a driving factor for the emergence of opinion polarization in online social media [281].

The contact pattern among the agents, sustaining the opinion formation, is described by a temporal network $A_{ij}(t)$, which changes in discrete time steps. Following empirical observations, we model the dynamics of $A_{ij}(t)$ by the activity-driven (AD) model [59, 194, 297, 298], see Ch. 2 for more information on the AD model. In short, each agent i is characterized by an activity $a_i \in [\varepsilon, 1]$, which corresponds to their propensity to contact m distinct random other agents in the system. We assume the activity distribution $F(a)$, from which the individual activities a_i are drawn, to follow a power law with exponent γ, $F(a) \sim a^{-\gamma}$, as measured empirically [59]. The parameter ε defines the lower cut-off for the activities to avoid the divergence of $F(a)$ close to $a = 0$. In the original formulation of the AD model, an active agent connects to m uniformly random sampled agents, and therefore the model is fully defined by the parameters (ε, γ, m). Here, we extend the AD model by homophily and assume that an agent preferentially contacts similar peers, where similarity is defined based on the opinion difference between two agents $|x_i - x_j|$. The probability p_{ij} that an active agent i will contact agent j is modeled as a decreasing function of their opinion distance,

$$p_{ij} = \frac{|x_i - x_j|^{-\beta}}{\sum_j |x_i - x_j|^{-\beta}}, \tag{7.3}$$

where β controls the power law decay of the connection probability with increasing opinion distance. It is important to note that the parameter β can contain different contributions that lead to overall homophilic interactions. This ranges from endogeneous effects, arising from intrinsic behaviors of homphilic agents ("birds of a feather flock together" [128]), to exogeneous ones due to e.g. recommender systems implemented on social media platforms [299].

7.2 Dynamical regimes

The following analysis focuses on a regime of fast-switching interactions, which is motivated by the highly dynamical and ever accelerating character of online social media environments [8, 85, 86, 87, 88, 89]. Furthermore, attitude change has been shown to be slow especially

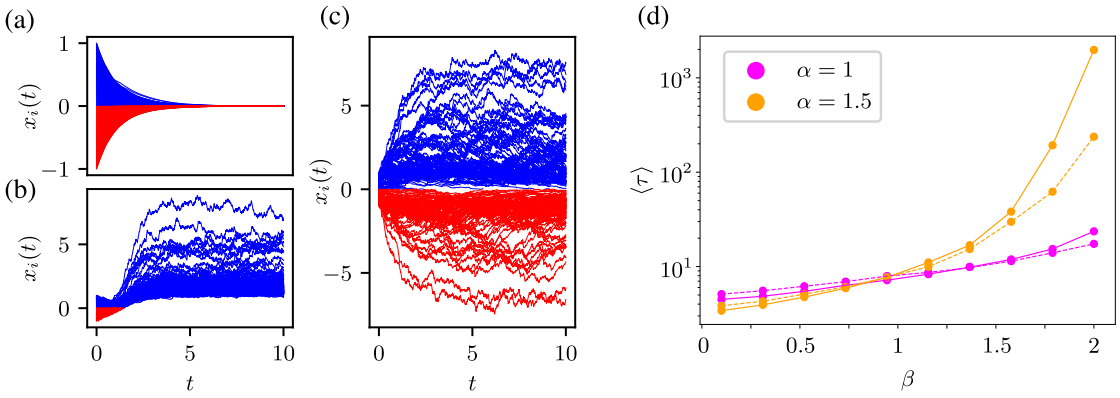

Figure 7.2: Panel (a)-(c): Different dynamical regimes of the radicalization model. (a) Convergence of the agents' opinions resulting in a neutral consensus ($\alpha = 0.05$, $\beta = 2$). (b) One-sided radicalization ($\alpha = 3$, $\beta = 0$). (c) Bi-polarized opinion state with two opposite opinion groups. The social interaction strength and the reciprocity were set to $K = 3$ and $r = 0.5$, respectively. Positive (negative) opinions with $\mathrm{sgn}(x_i) = 1$ ($\mathrm{sgn}(x_i) = -1$) are shown in blue (red). Note the different scales on y-axes. Panel (d): Mean lifetime of polarized opinion states such as the one depicted in panel (c). The lifetime strongly increases with the value of β (homophily). Each dot depicts the average of 1000 simulation runs for $N = 250$ and $dt = 0.05$. The colors (orange and magenta) correspond to different values of α, while we have set $K = 1$ in all runs. The dashed and solid lines show results for $r = 0.5$ and $r = 1$, respectively.

with respect to important personal issues [300]. These findings translate into a rather large time-scale separation between the network and the opinion dynamics with respect to controversial issues. Specifically, we numerically integrate Eq. (7.1) with $dt = 0.01$, and update the temporal network $A_{ij}(t)$ for each integration time step, as discussed in detail in App. B. We investigate the emerging opinion dynamics as a function of the social interaction strength K, the controversialness α, and the homophily exponent β, and fix the system size to $N = 1000$ agents. For each simulation run the initial opinions are uniformly spaced on the interval $x_i(0) \in [-1, 1]$, and, unless indicated differently, we set the AD parameters and the reciprocity r to $(\varepsilon = 0.01, \gamma = 2.1, m = 10)$ and $r = 0.5$, respectively. In App. B, we verify the robustness of the results for different reciprocity values.

Depending on the values of K, α and β, we identify three different dynamical regimes. In the case of both weak social interactions (low K) and small controversialness (low α), a *neutral consensus* arises, in which the opinions of all agents collectively converge to zero, due to insufficient social inputs. This is shown in panel (a) of Fig. 7.2. This state of global consensus is destabilized by larger values of K and/or α, which give rise to radicalization dynamics. In contrast to consensus states [panel (a)], where all agents converge to a neutral opinion, radicalization dynamics yield opinion distributions that are widely spread with some agents reaching opinions far outside the initial interval. Importantly, the dynamics and final states crucially depend on how the agents choose their interaction partners. In the absence of homophily ($\beta = 0$), as in the original AD model, all opinions will be directly absorbed by one of

the two opinion stances ($\mathrm{sgn}(x) = \pm 1$), as shown in panel (b). The introduction of homophily ($\beta > 0$) gives rise to a different dynamics: driven by repeated interactions with like-minded peers, agents reinforce their opinions. Homophily effectively stabilizes two opposed opinion camps (shown as blue and red trajectories), and leads to a segregation of the population into two groups. This situation is shown in Fig. 7.2(c), where only the value of β is increased in comparison to the case depicted in panel (b). Hence, for sufficiently large values of K, α and β a *polarized state* characterized by a bimodal opinion distribution emerges.

The observed polarized states are *metastable* and eventually turn into one-sided radicalized states, as the one shown in Fig. 7.2(b). Their lifetimes, however, increase at least exponentially with increasing values of the homophily exponent β, up to a point, where their destabilization becomes numerically inaccessible. The results of the corresponding numerical explorations are shown in Fig. 7.2(d), where we depict the super-exponential growth of the average lifetime $\langle\tau\rangle$ of polarized states [see panel (c) of Fig. 7.2] as a function of β, and for different values of the controversialness (α), as well as the reciprocity (r). While r only weakly affects the lifetime of polarized states, the level of controversialness does strongly so. Especially for high values of β, their lifetime significantly increases for slightly larger values of α. The qualitative explanation for the emergence of opinion polarization as a result of homophily is as follows. After their initialization, symmetrically distributed around the neutral consensus, each agent finds relatively fast its quasi-equilibrium, corresponding to a metastable opinion x. While in the absence of homophily ($\beta = 0$) agents' opinions rapidly relax to a global equilibrium on either side of the neutral consensus ($x = 0$), the introduction of homophily drastically changes this picture: agents with similar opinions interact strongly and form a metastable phase with well-defined opinions. By contrast, agents with different opinions hardly communicate, rendering the two phases practically non-interacting, and therefore long-living.

7.3 Transition to radicalization dynamics

To shed more light on the destabilization of the global consensus state, we study the model within a mean-field approximation. In the limit of fast-switching interactions, the adjacency matrix $A_{ij}(t)$ may be replaced by its time average $\langle A_{ij}(t)\rangle_t$. Thus, Eq. (7.1) yields

$$\dot{x}_i = -x_i + K \sum_{j=1}^{N} \langle A_{ij}(t)\rangle_t \tanh(\alpha x_j)\,. \tag{7.4}$$

The expression for $\langle A_{ij}(t)\rangle_t$ can be derived as follows. The total probability, that agent i will be influenced by agent j, has two contributions: (i) the probability that agent j contacts agent i (p_{ij}), and (ii) the probability that agent i contacts agent j *and* the link is reciprocated ($r\,p_{ji}$). Neglecting homophily ($\beta = 0$) the probability that agent i contacts agent j is simply $p_{ij} = m a_j / N$. Hence, summing both contributions we get

$$\langle A_{ij}(t)\rangle_t = \frac{m}{N}(a_j + r\,a_i)\,, \tag{7.5}$$

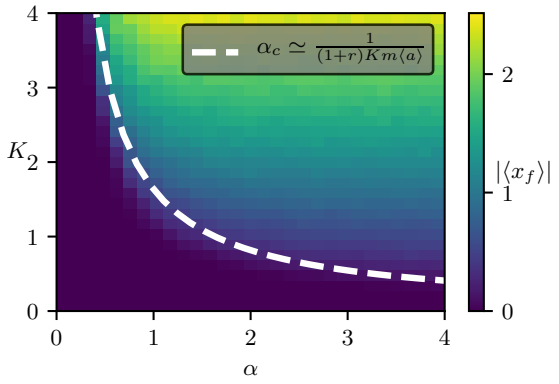

Figure 7.3: Transition between consensus and radicalization dynamics. Absolute values of the average final opinion $|\langle x_f\rangle|$, obtained from numerical simulations, in the K-α plane for $\beta = 0.5$ and $r = 0.5$. While for low values of K and/or α the system approaches consensus (dark purple region), the brighter areas correspond to radicalization dynamics for increasing values of K and/or α (color code). The transition is captured reasonably well by the mean-field approximation shown as white dashed line.

for the time averaged connectivity. The individual activities a_i are drawn from a power law distribution, with $F(a) = \frac{1-\gamma}{1-\epsilon^{1-\gamma}} a^{-\gamma}$. To obtain the average value of $\langle A_{ij}(t)\rangle_t$ with respect to all activities in the system we compute $\int_\epsilon^1 \langle A_{ij}(t)\rangle_t F(a)da = m(1+r)\langle a\rangle/N$, with $\langle a\rangle = \frac{1-\gamma}{2-\gamma}\frac{1-\epsilon^{2-\gamma}}{1-\epsilon^{1-\gamma}}$. Then we can write the mean-field version of Eq. (7.1) as

$$\dot{x}_i = -x_i + K\Lambda \sum_{j=1}^{N} \tanh(\alpha x_j), \tag{7.6}$$

where we defined $\Lambda = m(1+r)\langle a\rangle/N$. To evaluate the linear stability of the neutral consensus, we compute the Jacobian of Eq. (7.6) at $\mathbf{x} = \mathbf{0}$, which reads

$$\mathbb{J}(\mathbf{x}=\mathbf{0}) = \begin{bmatrix} -1 & K\Lambda\alpha & \dots & K\Lambda\alpha \\ K\Lambda\alpha & -1 & \dots & K\Lambda\alpha \\ \vdots & \vdots & \vdots & \vdots \\ K\Lambda\alpha & K\Lambda\alpha & \dots & -1 \end{bmatrix}, \tag{7.7}$$

where all off-diagonal elements are equal to $K\Lambda\alpha$. The stability of consensus is determined by the largest eigenvalue $\tilde{\lambda}$ of $\mathbb{J}(\mathbf{0})$, which reads

$$\tilde{\lambda} = (N-1)K\alpha\Lambda - 1 = \frac{(N-1)}{N} K\alpha m(1+r)\langle a\rangle - 1. \tag{7.8}$$

For $\tilde{\lambda} < 0$ the neutral consensus is stable and no radicalization dynamics takes place. Thus, $\tilde{\lambda} = 0$ defines the transition line between the consensus and radicalized phase. Solving Eq. (7.8)

for α yields, in the thermodyamic limit with $\lim_{N\to\infty}\left(\frac{N-1}{N}\right)=1$, the expression for the critical controversialness

$$\alpha_c \simeq \frac{1}{(1+r)Km\langle a\rangle}, \tag{7.9}$$

for which the neutral consensus becomes unstable and radicalized states emerge. The critical α_c is inversely proportional to the product of the social interaction strength K, the number of contacts per active agent m, the average activity $\langle a\rangle$, and the term $(1+r)$ accounting for the reciprocity of social influence. In Fig. 7.3, we compare this mean-field result to numerical simulations of the full system, where Eq. (7.9) is depicted as white dashed line in the K-α plane. The colors encode the absolute value of the final average opinion, $|\langle x_f\rangle| \equiv |N^{-1}\sum_i x_i(t_{\text{final}})|$ for a specific combination of K and α. For long simulation times, the value of $|\langle x_f\rangle|$ identifies the transition between regions of consensus (dark purple) and regions of radicalized states (color coded blue to yellow). While consensus is obtained for small values of K and/or α, radicalization arises for increased social interaction strengths and/or large values of controversialness. Note that, although we neglected homophily in the derivation of Eq. (7.9), the mean-field approximation still captures the consensus-radicalization transition for moderate values of β.

7.4 Polarized debates on Twitter

We now contrast the model results to three data sets of polarized debates on *Twitter*, which were collected by K. Garimella and co-workers [301] and previously analyzed in Ref. [71]. The Twitter data contains tweets on specific topics, which are known to be politically controversial: *gun control*, *Obamacare*, and *abortion*. The data sets are built along two main features: (i) the political leaning of users (opinions), and (ii) their social interaction network [71]. To approximate users' opinions towards the different topics the following procedure was applied. Tweets were collected during specific events that trigger a collective interest in a given topic. For each topic, the time span of data collection was one week – three days before and three days after the event. To assign a tweet to a specific topic (*gun control*, *Obamacare*, and *abortion*), an established list of keywords was used [302]; whenever one of such keywords was mentioned in a tweet, it was assigned to the associated topic. The political leaning score of each tweet was then obtained via links to news organizations contained in the tweet, where each news organization (e.g. *nytimes.com*) was assumed to have an associated leaning score [72]. More specifically, each news source was classified by a score which takes values between 0 and 1, where a value of 0 (1) indicates that the news source has a very conservative (liberal) leaning. For instance, if a tweet contained the link to a New York Times article, the tweet gets assigned the leaning score of *nytimes.com*. From this "ground truth" the political leaning of a user was obtained as the average leaning of all tweets of the user during the time span of data collection. For a clearer comparability with the model we map the interval $[0,1]$ of leanings to $x_i \in [-1,1]$ (similar to positive and negative opinions x_i in the model). In Ref. [71], the social interaction network for the data sets was reconstructed as the follower network among users: a directed link (i,j) indicates that user i follows user j. In the following, we contrast the Twitter data with our model using a reciprocity value of $r=0.65$, which is close to the average

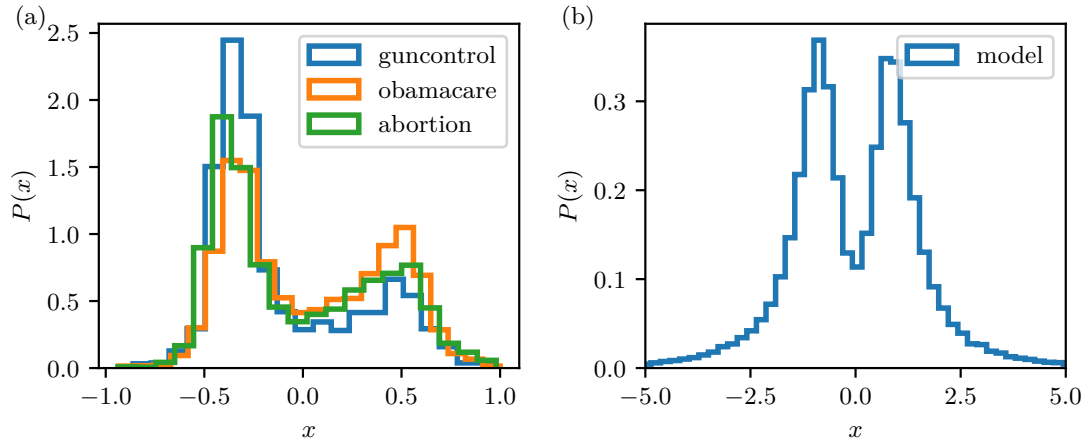

Figure 7.4: Normalized opinion distributions obtained from the three investigated Twitter data sets (a) and by simulating the model with parameters $K = 3$, $\alpha = 3$, $\beta = 1$, $r = 0.65$. With these parameters the model enters a polarized state, characterized by a bi-modal opinion distribution, similarly found for the Twitter data sets of polarized debates.

reciprocity value of the three data sets ($r \simeq 0.64$). See App. B, for further details on the utilized Twitter data.

Opinion bi-polarization. First, we consider the opinion distributions $P(x)$ obtained for the Twitter data, which are depicted in panel (a) of Fig. 7.4. Clearly, the distributions of all three data sets have a pronounced bimodal shape. This characteristic feature has been found for polarized debates on various controversial issues, including data from other social media platforms, where the method to infer individual opinions may differ, such as likes of Facebook pages [16] or up votes of Youtube videos [303]. In the polarized regime, i.e. for $\alpha > \alpha_c$ and sufficient levels of homophily β, the bimodal shape of the opinion distribution is qualitatively reproduced by the model, as depicted in panel (b). While in this regime bimodal opinion distributions are a stable result for a wide range of model parameters, the width of the gap between the two peaks strongly depends on $F(a)$. More specifically, the gap generally widens and becomes much more pronounced for higher activity cut-offs, or larger mean activities. This is illustrated in Fig. B.2 of App. B for two cases of delta distributed activities, i.e., $F(a) = \delta(a - a_c)$, where each agent has activity a_c.

Association between activity and opinion. Figure 7.5 shows another interesting feature found in different datasets of polarized debates, namely a strong association between the engagement of users in the discussion and their convictions. In particular this means: more-active users tend to show more-extreme opinions. In the case of the Twitter data analyzed here, we asses the activity a of a user as the fraction of tweets containing links to news organizations, as an approximation of their political engagement. In Fig. 7.5(a), we plot the average engagement (or activity) of users against their inferred opinions x. For all three considered topics the

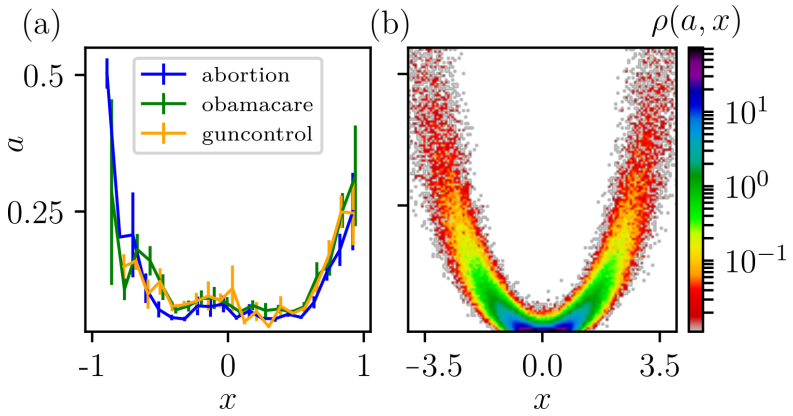

Figure 7.5: Association between activity and opinion. (a) Average binned activity $\langle a \rangle$ plotted against the political leaning x of users, for three empirical data sets. (b) Activity-opinion density plot of 10^3 polarized opinion states with $K = 2$, $\alpha = 2$, $\beta = 1$, and $r = 0.65$. The colors encode the density value of $\rho(x, a)$ normalized with respect to the system size N.

engagement rises considerably towards the boundaries of the opinion space. Similar relations have also been found for differently defined user activities and opinions, including shares of political content on Facebook [72], or tweet rates of users based on the hashtags they use [85]. In the polarized regime, the characteristic U-shaped relation is qualitatively reproduced by the model. In Fig. 7.5(b), we show the corresponding results as a density plot encompassing the polarized states of 10^3 simulation runs. The finding suggests the following: while most users have a rather low activity and convictions close to the neutral consensus, some very active agents develop more-extreme opinions. Due multiple interaction with like-minded peers highly active users reinforce their opinions, which are driven away from consensus and towards more-extreme convictions.

Echo chambers. Finally, we check if the model captures echo chambers with a similar signature as observed in the Twitter data. Following Ref. [85], we define echo chambers in terms of the correspondence between the distribution of opinions $P(x)$ and the underlying social influence network: users are more likely to be connected to peers with similar opinions, which may foster information exchange among like-minded individuals. On a network level, this correspondence translates into a correlation between the opinion of user i, x_i, and the average opinion of their nearest neighbors $\langle x_i \rangle^{NN}$. For the Twitter data we consider the underlying follower-followee network: $\langle x_i \rangle^{NN}$ is computed as the average opinion of users j followed by user i. In the model, there is no underlying static network topology and social interactions evolve dynamically. Therefore, the average opinion of agent i's neighbors is computed based on an aggregated interaction network A_{ij}^{agg}, i.e., we have $\langle x_i \rangle^{NN} = k_i^{-1} \sum_j A_{ij}^{\text{agg}} x_j$. For A_{ij}^{agg} we take into account 45 network snapshots and start the aggregation once the system is in a polarized state.

Figure 7.6 shows the correlation between agents' opinions and the opinions of their neigh-

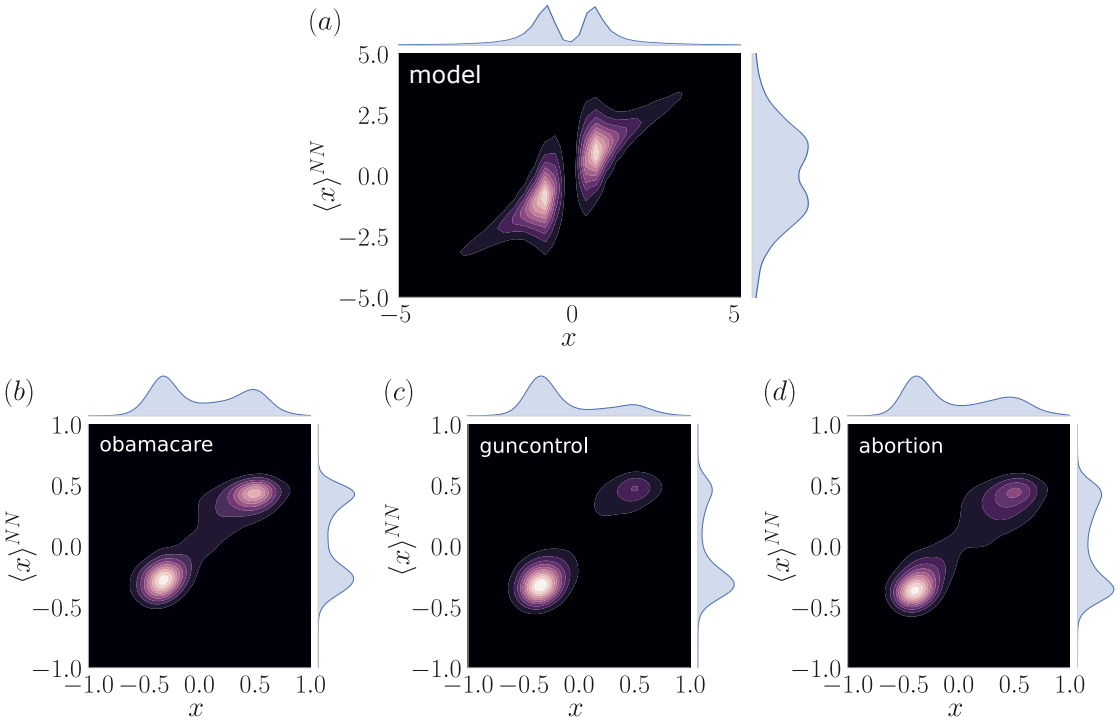

Figure 7.6: Echo chambers depicted as contour maps for the average opinion of the nearest neighbors $\langle x \rangle^{\mathrm{NN}}$ plotted against a user's opinion x for 200 simulations of the model for $K = 2.5$, $\alpha = 4.5$, $\beta = 2$, and $r = 0.65$ (a) and the three data sets (b)-(d). Colors encode the density of users, where the brightness increases with the density. The marginal distributions of opinions $P(x)$ and average nearest neighbor opinions $P_{\mathrm{NN}}(x)$ are plotted on the x and y axis, respectively.

bors in terms of contour maps, both for the model [panel (a)] and for the three empirical datasets, see panels (b)-(d). The color code represents the number of users in the phase space $(x, \langle x_i \rangle^{NN})$, where the brightness corresponds to the density of users (or agents): the lighter the area, the larger the density of users (or agents). In all four plots, there are strong correlations between the opinion of a user (agent) and the average opinion of their neighbors. Both the results of our model and the ones of the Twitter data show two bright "bubbles" of high user densities. Each of these bubbles identifies an echo chambers, where users (agents) interact with like-minded peers. The bubbles are (more or less) disconnected suggesting that they correspond to opposite and rather isolated opinion groups. In our model, the existence of pronounced echo chambers indicates a feedback loop between homophily and social reinforcement: while homphily yields contacts between like-minded agents, social influence among similar agents reinforces their opinions rendering interactions to dissimilar peers increasingly unlikely.

7.5 Chapter summary and discussion

Abelson's question on how opinion cleavages may emerge and persist over time has been tackled by previous modeling efforts, some of which were discussed in Ch. 6. Here, we introduced a minimal model based on a different set of assumptions. A simple social reinforcement mechanism was assumed, which is inspired by persuasive-arguments theory and group polarization: agents' convictions are increased upon the interactions with like-minded peers. Notably, the model does neither assume a fixed network topology nor a sharp confidence bound. In contrast, agents' opinions and the temporal network of social interactions co-evolve based on homophily. We investigated the model, both numerically and analytically, and found that it reproduces stylized facts of empirical data about controversial debates on social media, such as opinion-activity associations, polarized opinion distributions, and the formation of echo chambers.

It is important to highlight that the comparison of our model with empirical Twitter data has major limitations. First, we could not *observe* the dynamics of opinion polarization and echo chambers in the Twitter data. Instead, we compared the stationary states of the model with a more or less instantaneous snapshot of the processes on Twitter. As a future work, it would be very interesting to check if the dynamics proposed by the model indeed captures the dynamics of "real" radicalization and echo chambers with a focus on their time evolution. Another limitation concerns the reliability of opinion inference. We did not ask people's opinions on different topics by e.g. a survey, but rather determined them indirectly with the help of a simple heuristics (links to news sites within tweets). In the conclusions of this thesis, we will discuss a recent proposal that addresses similar issues [276].

In the remainder of this chapter, we contrast our model with previous and especially related modeling efforts in the field. In particular, different social reinforcement mechanisms (in some cases combined with homophily) have been found to give rise to persistent heterogeneous

and polarized opinion states [54, 80, 121]. Persuasive-arguments theory (PAT) was previously implemented in a model of opinion dynamics in a very direct way [121], where agents explicitly exchange arguments. The framework was termed argument-communication theory of bi-polarization (ACTB). Slightly simplified, ACTB can be summarized as follows. Each agent i holds a vector of binary arguments, where each argument is either in favor $(+1)$ or against (-1) a certain topic. The time-dependent opinion of an agent is then given as the mean of all their current arguments. Social influence between agents is assumed to take place in discrete time steps and it is mediated via argument exchange: if there is a contact between agents i and j, the opinion of agent i is updated, such that a randomly chosen argument, held by agent j, is adopted by agent i. This process may have two qualitatively different outcomes with regard to the change of agents' opinions. On the one hand, agents' opinions may become more moderate due to an exchange of opposed arguments. On the other hand, social influence may lead to more extreme opinions if interactions take place among agents with similar sets of arguments. Similar to our model, ACTB gives rise to opinion bi-polarization in the case of homophily. Futhermore, it was found that the level of homophily has important implications on the collective opinion dynamics induced by ACTB: more simulation runs resulted in bi-polarized states for larger values of homophily. This is in line with our findings on one-sided radicalized states in the case of $(\beta = 0)$ and the lifetimes of polarized states increasing exponentially as a function of β. Although ACTB models the process of argument exchange much more realistically, our approach has some advantages with respect to mathematical tractability. For instance, within a mean-field theory we could precisely determine the factors which de-stabilize consensus. Furthermore, as we will see in Ch. 8, it is straightforward to generalize our model to multiple dimensions, an endeavour which seems much harder in the case of ACTB.

Next, we compare our model to an approach based on an information accumulation system (IAS) [80] in a population of agents whose influence network is characterized by two strongly pronounced communities. The overarching question addressed is the following: under which conditions do both communities reach consensus or maintain a state of opinion diversity? More specifically, it is asked how strong can be the communication, or "diffusivity", between the agents, such that no global consensus is established. Due to the pronounced community structure it is distinguished between communication within a community (intra-diffusivity) and between the communities (inter-diffusivity). In the IAS, an agent accumulates information from other agents, which is similar to the social input term in Eq.(7.1). Furthermore, the finite memory of agents is modeled in a similar manner as in our model, such that agents loose information, or their opinion towards a certain topic, if there are no inputs from other agents. Importantly, in contrast to classical assimilation models (Ch. 3), and similar to our radicalization model, the emergence of consensus is due to the lack of social inputs from other agents and can thus be interpreted as a relaxation process to the neutral state. Starting from a situation in which agents in different communities have opposite opinion stances (i.e., different signs of opinions) the IAS model gives rise to three different kinds of states: (i) states in which all agents have a neutral opinion, (ii) states in which agents in one community

have adopted the opinions of the agents in the other community, and (iii) polarized states in which both opinions co-exist. It is shown that the neutral state is reached for both low intra- and inter-community diffusivity. This is analogous to situations of consensus in our radicalization model. As Eq. (7.9) suggests, consensus emerges (given a fixed value of α) in the case of low $K, m, \langle a \rangle$ and $(1+r)$, i.e., for parameters values, which limit the amount and strength of communication (or social influence) among agents. For intermediate values of both the inter- and the intra-community diffusivity only one opinion survives, which gives rise to a system-wide consensus, whose value depends on the initial conditions. This is in line with one-sided radicalized states (see Fig. 7.2(b)), where the neutral consensus is de-stabilized due to increased levels of social influence among agents and high values of controversialness. Finally, the co-existence of two opinions in the IAS model is possible if the intra-community diffusivity strongly exceeds the inter-community diffusivity. Interestingly, in our model this corresponds to a situation of sufficient values of K and α, and high levels of homophily. For large values of β (high homophily), more interactions take place between like-minded agents, effectively forming two opposed communities with opposite opinion stances. The example of the IAS model illustrates how social reinforcement may give rise to opinion polarization on a fixed network structure, if there are pronounced communities. The effects of high homophily (in our model) can therefore be partly understood in terms of an even simpler framework, where communities are not formed but are already present.

Previous models have also been proposed to explain the emergence of echo chambers. In Ref. [304], the authors studied a model, in which agents are characterized by a continuous opinion and perform a random walk in a two-dimensional physical space. Agents interact such that they adapt their opinions to become more similar to the peers in their physical vicinity, modeled as an assimilative dynamics. If agents are close to other agents (in physical space) which have similar opinions, they tend to remain in such places. This mechanism creates a feedback loop, leading to the emergence of groups comprised of with similar opinions, i.e., the echo chambers. A similar model, based on assimilative dynamics and confidence bounds, was proposed in Ref. [89]. The model investigates how political campaigns, in combination with social peer-influence, determine the collective opinion dynamics with regard to voting behavior. It is found that intermediate levels of open-mindedness of agents (essentially their confidence bounds) may lead to bi-polarization and the formation of two opposed echo chambers for two antagonistic political campaigns. In our work, such external effects of e.g. political campaigns have been disregarded. However, in a recent preprint [305], which builds on our model, the authors investigate the effects of an additional noise term in Eq. (7.1), which models a simple nudging intervention [306]. In contrast to political campaigns, which usually tend to promote bi-partisanship, the introduced nudge aims to lower opinion polarization and to mitigate the emergence of echo chambers. Along the same lines, it would be possible to extend Eq. (7.1) by divisive factors to investigate the effects of e.g. opposed political campaigning on the stability of consensus and the emergence of radicalization.

8 Emergence of Polarized Ideological States in Multi-dimensional Topic Spaces

This chapter is mainly based on the publication

Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov and Michele Starnini.
Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces.
Phys. Rev. X 11 (2021): 011012.

We thank APS for the kind permission to reuse the contained figures for this dissertation.

Most previous models developed to explain opinion cleavages were formulated in one dimension. More specifically, it was assumed that agents hold and express opinions only with respect to a single topic. However, this assumption essentially implies that opinions with respect to different topics form independently. In communication ecosystems of modern societies with an ever growing connectedness and increased information flows between individuals[8, 307], as well as online social networks where topics are often discussed simultaneously this assumption should be questioned. The hypothesis about interdependent processes of opinion formation towards different topics is also suggested by empirical opinion data. On top of opinion polarization, various studies have revealed another striking feature of opinion distributions: issue alignment [116, 117, 118],which comes to light if the opinions towards multiple topics are considered simultaneously. Issue alignment implies that individuals are more likely to have a certain combination of opinion stances (in favor/against) than others. We refer to such states as ideological opinion states, which can be understood in terms of an intuitive example. Let us consider opinions towards the rights of transgender people, and opinions towards the rights of same-sex couples. Here, most people will presumably split into two groups: those who deny many rights to both minority groups, and those who support them, while the mixed positions would rarely occur. Indeed, those two topics are highly related, and therefore it does not come as a big surprise if public opinions towards them are correlated. However, as suggested by empirical data of the 2016 American National Survey (ANES), significant opinion correlations can also be observed for rather unrelated issues. Which underlying mechanism

might drive the emergence of such ideological states?

To approach this question, we generalize the one-dimensional radicalization model (Ch. 7) to multiple dimensions. Crucially, we assume that the underlying opinion space is spanned by the topics which are discussed simultaneously. This assumption is inspired by vector space models, where text documents are represented in an underlying space, whose basis is formed by the terms used in the documents [308]. In the regime of large homophily, the model gives rise to three qualitatively different dynamical regimes, which lead to opinion states that can be observed in two-dimensional opinion data: (i) the convergence to a global consensus, (ii) the polarization of non-correlated opinions, and (iii) the opinion polarization with issue alignment, or polarized ideological states. Interestingly, ideological states emerge from uncorrelated polarization by relaxing the assumption of orthogonal topics. In this case, topics may exhibit a certain degree of overlap, which has an intuitive interpretation in terms of persuasive-arguments theory [30, 31]: the overlap between two topics may represent a common set of arguments, which simultaneously supports or rejects a certain stance (in favor/against) of both considered topics. Hence, large overlaps characterize closely related topic pairs, where there are multiple related arguments. If only a few arguments concern both topics simultaneously, the overlap will be small. We study the transitions between the dynamical regimes (i)-(iii) as functions of the controversialness of the topics discussed and their pairwise overlaps and find that polarized ideological states with high opinion correlations may indeed emerge also for very small topic overlaps. Contrasted with opinion polls of the 2016 American National Election Surveys (ANES) [90], we find that the model qualitatively reproduces crucial two-dimensional opinion states found in the data. It also offers possible explanations for the observed issue alignment between topics which are rather unrelated but sufficiently controversial.

This chapter is structured as follows. In Sec. 8.1, we formally generalize the radicalization model to multiple dimensions. In Sec. 8.2 and Sec. 8.3, we discuss the emerging opinion dynamics in the full model and in a mean-field approximation for a two topics, respectively. In Sec. 8.4, we consider a three-dimensional topic space. In Sec. 8.5, we investigate the resulting opinion segregation in the network of social interactions. Finally, in Sec. 8.6, we contrast the model to empirical data drawn from the American National Election Studies. The chapter is concluded in Sec. 8.7, where we discuss our results with respect to previous multi-dimensional models of opinion dynamics.

8.1 Radicalization model for multiple topics

In contrast to the one-dimensional model (Ch. 7), where agent i is characterized by an opinion scalar $x_i \in \mathbb{R}$, in the following, each agent holds opinions towards T distinct topics. Those opinions are stored in the time-varying opinion vector $\mathbf{x}_i(t) = (x_i^{(1)}(t), x_i^{(2)}(t), \ldots, x_i^{(T-1)}(t), x_i^{(T)}(t)) \in \mathbb{R}^N$. More specifically, the component $x_i^{(\nu)}(t)$ is the opinion of agent i towards topic ν at time t, where we will omit the time-dependence for brevity. The definitions of both the opinion

stance and the conviction of an agent carry over from the one-dimensional model, but are now defined separately for each topic. More specifically, for a given topic ν the sign of $x_i^{(\nu)}$, $\mathrm{sgn}(x_i^{(\nu)})$, defines the qualitative stance of agent i towards this topic, and the absolute value $x_i^{(\nu)}$, $|x_i^{(\nu)}|$, quantifies the agent's conviction towards topic ν. The opinion vector $\mathbf{x}_i$ defines the position of agent i's opinion in the T-dimensional Euclidean topic space $\mathscr{T}$, and can be written as $\mathbf{x}_i = \sum_{\nu=1}^{T} x_i^{(\nu)} \mathbf{e}^{(\nu)}$, where $\{x_i^{(\nu)}\}$ and $\{\mathbf{e}^{(\nu)}\}$ define the opinion coordinates of agent i and the basis vectors of $\mathscr{T}$, respectively. Note that the set of $\{\mathbf{e}^{(\nu)}\}$ is linearly independent, but not necessarily orthogonal.

As in the one-dimensional case, we assume that the social interactions, sustaining the opinion dynamics, evolve in time according to the activity-driven (AD) model [59], where $A_{ij}(t) = 1$ if agents i and j are connected at time t, and $A_{ij}(t) = 0$ otherwise[1]. Equation (7.1) is generalized to multiple dimensions in the most straightforward way, yielding

$$\dot{x}_i^{(\nu)} = -x_i^{(\nu)} + K \sum_j A_{ij}(t) \tanh\left(\alpha \left[\Phi \mathbf{x}_j\right]^{(\nu)}\right), \tag{8.1}$$

where, correspondingly, the parameters K and α denote the social interaction strength among agents and the controversialness of topics, respectively. Note the introduction of the topic overlap matrix Φ, which will be defined and explained a couple of lines below.

According to Eq. (8.1), the opinion of agent i towards topic ν, $x_i^{(\nu)}$, evolves depending on the aggregated inputs from their neighbors, encoded in the time-varying adjacency matrix $A_{ij}(t)$. Importantly, in the multi-dimensional case, the social input from agent j contributing to the change of $x_i^{(\nu)}$, $\left[\Phi \mathbf{x}_j\right]^{(\nu)}$, is generally a function of the entire opinion vector of agent j. As in Ch. 7, we choose as non-linear influence function the $\tanh(\cdot)$, which is tuned by the controversialness α. Note, however, in Eq. (8.1), we assume an overall controversialness value which holds for all topics. Below, in Sec. 8.2, we discuss a more general case, where we consider different values of α for each topic.

The crucial ingredient of the multi-dimensional model is the symmetric topic overlap matrix Φ. It encodes the relations among the T topics with implications for the macroscopic opinion dynamics. If the matrix element $\Phi_{u\nu}$ is different from zero, agents' opinions on topic u can influence their peers' opinions on topic ν, and vice versa. Geometrically, the matrix Φ may be interpreted in the underlying topic space as follows. The element $\Phi_{u\nu}$ corresponds to the scalar product of topics u and ν, $\Phi_{u\nu} = \mathbf{e}^{(u)} \cdot \mathbf{e}^{(\nu)} = \cos(\delta_{u\nu})$, where $\delta_{u\nu}$ denotes the angle between topics u and ν, as depicted in Fig. 8.1 for $T = 2$. With respect to our introductory example, $\cos(\delta_{u\nu})$ corresponds to the topical overlap between topic u ("rights of transgender people") and topic ν ("rights of same-sex couples"). Those pairwise overlaps also appear in the scalar product between two opinion vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ in the topic space $\mathscr{T}$ as

$$\mathbf{x}_i \cdot \mathbf{x}_j = \mathbf{x}_i^T \Phi \mathbf{x}_j = \sum_{\nu,z} x_i^{\nu} x_j^{z} \cos(\delta_{\nu z}), \tag{8.2}$$

[1] In contrast to the one-dimensional radicalization model, we assume a fully reciprocal network, i.e. $r = 1$.

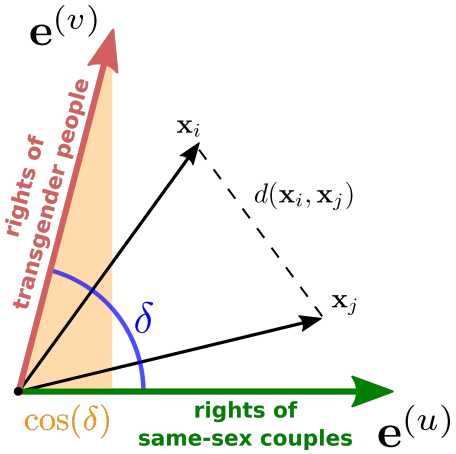

Figure 8.1: Illustration of two non-orthogonal topics forming the basis of a two-dimensional topic space $\mathscr{T}$. For two-dimensional topic spaces the non-orthogonal, normalized basis is entirely defined by the angle δ. Geometrically, $\cos(\delta)$ is the overlap between the two basis vectors, which is interpreted as a topical overlap, e.g. the *rights of same-sex couples* ($\mathbf{e}^{(u)}$) and the *rights of transgender people* ($\mathbf{e}^{(v)}$). The opinion distance between two agents with opinions $\mathbf{x}_i$ and $\mathbf{x}_j$, $d(\mathbf{x}_i, \mathbf{x}_j)$, is computed involving the scalar product, as defined in Eq. (8.2).

which involves the topic overlap matrix Φ. Note that it always holds that $\Phi_{uu} = 1$, such that the matrix Φ reduces to the unit matrix if all topics are orthogonal ($\Phi_{uv} = 0$). In that case, Eqs. (8.1) decouple with respect to topics, recovering the one-dimensional dynamics described by Eq. (7.1).

Due to the generalization to multiple dimensions, the implementation of homophily has to be adapted as well. The probability that an active agent i will contact a peer j is now modeled as

$$p_{ij} = \frac{d(\mathbf{x}_i, \mathbf{x}_j)^{-\beta}}{\sum_j d(\mathbf{x}_i, \mathbf{x}_j)^{-\beta}}, \tag{8.3}$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between two opinion vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, which is computed based on the scalar product defined in Eq. (8.2). As in the one-dimensional case, β controls the power law decay of the connection probability p_{ij} with increasing opinion distance.

Upon the interaction of agents i and j, all opinions of agent j may influence the opinions of agent i. Generally, we distinguish two cases: (i) $\Phi = \mathbb{1}$, and (ii) $\Phi \neq \mathbb{1}$. In the first case of orthogonal topics, only opinions towards the same topics influence each other. If both stances on a topic v are equal $\left(\text{sgn}(x_i^{(v)}) = \text{sgn}(x_j^{(v)})\right)$, agents will increase their current convictions with respect to topic v. By contrast, for $\text{sgn}(x_i^{(v)}) \neq \text{sgn}(x_j^{(v)})$, they will tend to converge towards the neutral stance. Importantly, for non-orthogonal topics ($\Phi \neq \mathbb{1}$) with $\cos(\delta_{uv}) \neq 0$, the opinion of agent j on topic v, $x_j^{(v)}$, will influence the opinion of agent i on topic u, $x_i^{(u)}$.

8.2 Emergence of consensus, polarization, and ideological states

Building on the results of the radicalization model in one dimension ($T = 1$) we now explore the effects induced by multiple topics ($T > 1$) and their pairwise overlaps. We consider a system of $N = 1000$ agents, set the parameters of the AD model to $(\varepsilon, \gamma, m) = (0.01, 2.1, 10)$, and focus on a regime of both strong social influence ($K = 3$) and strong homophily ($\beta = 3$). As in Ch. 7, we consider fast-switching interactions leading to a time-scale separation between the network and the opinion dynamics of a factor of 100. For details on the numerical simulation of the model including the generation of the temporal network $A_{ij}(t)$, see App. B.

We investigate the emergence of different opinion states depending on the controversialness α and the pairwise topic overlaps $\cos(\delta_{uv})$. As observed for the one-dimensional model, the polarized opinion states are not stable for $t \to \infty$, due to the stochastic dynamics of the underlying interaction network. However, for sufficiently high values of homophily (β) they have been found to be metastable and numerically indistiguishable from stable states, see Fig 7.2(d). The same holds true for multiple topic dimensions.

For the sake of simplicity and convenient illustrations, we will first disuss the model for two topics u and v. For $T = 2$, Eqs. (8.1) yield

$$\begin{aligned} \dot{x}_i^{(u)} &= -x_i^{(u)} + K \sum_j A_{ij}(t) \tanh\left[\alpha(x_j^{(u)} + \cos(\delta) x_j^{(v)})\right] \\ \dot{x}_i^{(v)} &= -x_i^{(v)} + K \sum_j A_{ij}(t) \tanh\left[\alpha(\cos(\delta) x_j^{(u)} + x_j^{(v)})\right], \end{aligned} \tag{8.4}$$

where Φ is entirely defined by a single angle $\delta_{uv} \equiv \delta$, with $\cos(\delta)$ defining the topic overlap between topics u and v.

In Fig. 8.2, we show the three dynamical regimes of the model. They crucially depend on both the controversialness α and the topic overlap $\cos(\delta)$. Grey lines depict the opinion trajectories of single agents, and their steady state positions are shown as colored dots. For the sake of visualization we use polar coordinates (r, φ), with $x^{(u)} = r\cos(\varphi)$ and $x^{(v)} = r\sin(\varphi)$, where r corresponds to the overall conviction of an agent that is colored according to the opinion angle φ. In panel (a) of Fig. 8.2, the topics are not controversial and agents reach a global consensus. Starting from a normally distributed two-dimensional initial state, all agents rapidly loose their convictions and converge to the neutral stance, i.e. , $||\mathbf{x}_i(t \to \infty)|| = 0$. As in one dimension, this fast decay is the result of the decay terms $(-x_i^{(u)}, -x_i^{(v)})$ in Eqs. (8.4) and insufficient social influence among agents. The corresponding opinion distributions, $P_u(x)$ and $P_v(x)$, which are depicted on the marginals of Fig. 8.2(a), are peaked around $x = 0$. By contrast, if topics are controversial (for larger values α) the dynamics is dominated by reinforcing social influence among agents, which leads to a destabilization of consensus. Those cases are shown in panels (b) and (c) of Fig. 8.2. The opinions do not converge, and instead, are widely spread and may reach convictions that are much larger than those of the initial opinion state. To observe such polarized states homophily is a necessary condition, as we have already discussed for the

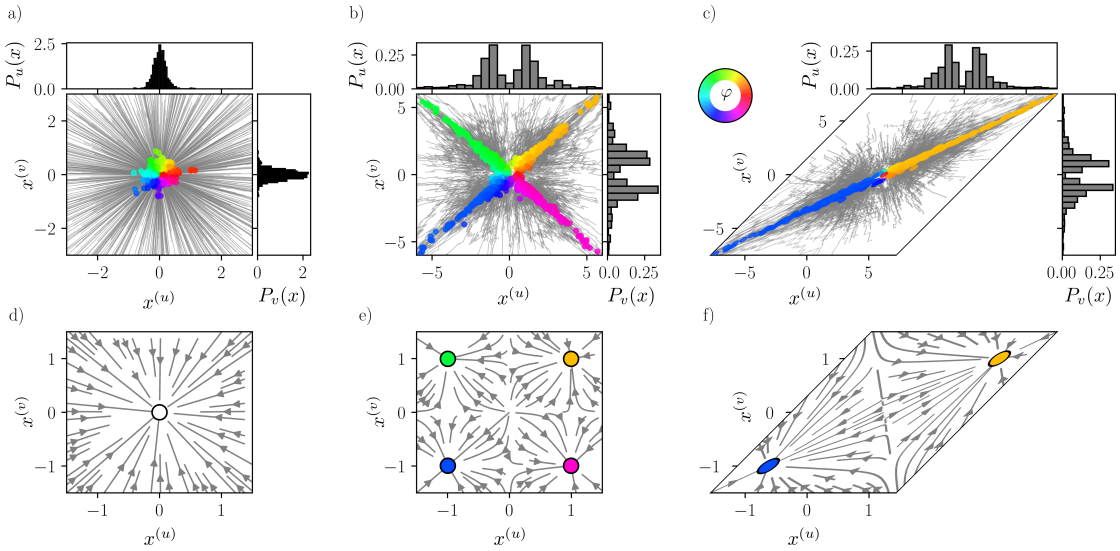

Figure 8.2: Consensus, uncorrelated polarization and polarized ideological states in a two-dimensional topic space. We show opinion evolutions from numerical simulations of the full stochastic model [panels (a)-(c)] and the corresponding deterministic attractors [panels (d)-(f)], obtained from a mean-field approximation, with identical values of α and δ for each pair [(a), (d)], [(b), (e)], and [(c), (f)]. The opinion trajectories of individual agents are depicted as grey lines and the final opinions $\mathbf{x}_i$ are colored according to the opinion angle φ_i. The agents approach a global consensus if topics are not controversial, as for low $\alpha = 0.05$ (a), while opinion polarization emerges for controversial topics ($\alpha = 3$), depicted in panels (b) and (c). The variances of marginal distributions $P_u(x)$ and $P_v(x)$, $\sigma_u^2(x)$ and $\sigma_v^2(x)$, reflect the degree of polarization. The variances are low for consensus ($\sigma_u^2(x) = 0.04$, $\sigma_v^2 = 0.035$) and high for polarized states, i.e. $\sigma_u^2(x) = 7.27$, $\sigma_v^2 = 7.17$ in panel (b) and $\sigma_u^2(x) = 11.22$, $\sigma_v^2 = 11.2$ in panel (c). If topics do not overlap for $\delta = \pi/2$, all opinion combinations appear in the consensus (a) and in the uncorrelated polarized state (b), resulting in low correlation values of $\rho(x^{(u)}, x^{(v)}) = 0.01$ (a) and $\rho(x^{(u)}, x^{(v)}) = 0.024$ (b). Instead, for overlapping topics with e.g. $\delta = \pi/4$, as shown by the angle between the x and y axis in panels (c) and (f), opinions become strongly correlated ($\rho(x^{(u)}, x^{(v)}) \simeq 1$) and polarized ideological states emerge.

one-dimensional model. In App. B, we demonstrate the important role of homophily also for opinion polarization in higher dimensional topic spaces by simulating Eqs. (8.4) for low, intermediate, and high values of β.

In the following, we will examine the critical role of the topic overlap $\cos(\delta)$. If topics do not overlap ($\cos(\delta) = 0$), Eqs. (8.4) decouple with respect to topics u and v. Accordingly, the opinion dynamics with respect to each topic is independent and governed by the one-dimensional dynamics (Ch. 7). The corresponding polarized state is shown in Fig. 8.2(b). It is characterized by bimodal opinion distributions, $P_u(x)$ and $P_v(x)$, with respect to both topics u and v, as depicted on the marginals of the plot. In contrast to the consensus state, the variances of both distributions, $\sigma_u^2(x)$ and $\sigma_v^2(x)$, are large, as reported in the caption of Fig. 8.2. More importantly, for orthogonal topics, all possible combinations of opinion stances are realized in the polarized steady state, i.e. $[\mathrm{sgn}(x_i^{(u)}), \mathrm{sgn}(x_i^{(v)})] \in \{(-,+),(+,+),(-,-),(+,-)\}$. This translates into a low opinion correlation $\rho(x^{(u)}, x^{(v)})$, similar to the one of the consensus state, see caption of Fig. 8.2.

However, if topics overlap ($\cos(\delta) > 0$), i.e. , if the underlying space $\mathcal{T}$ is spanned by non-orthogonal topics, the situation is drastically different. In this case, opinions with respect to different topics can influence each other. In panel (c), we show the dynamics for $\delta = \pi/4$, which corresponds to $\cos(\delta) = 1/\sqrt{2}$. In contrast to the orthogonal case ($\delta = \pi/2$), not all combinations of opinion stances are realized. Instead, the dynamics selects opinion combinations, where agents share the same stance on both topics, i.e. $[\mathrm{sgn}(x_i^{(u)}), \mathrm{sgn}(x_i^{(v)})] \in \{(-,-),(+,+)\}$, while other combination gradually disappear as the system approaches the steady state. The marginal opinion distributions are again bimodal, but opinions with respect to different topics are highly correlated, with $\rho(x^{(u)}, x^{(v)}) \simeq 1$. We call this type of state with $\sigma_u^2(x), \sigma_v^2(x) \gg 1$ and high correlations $\rho(x^{(u)}, x^{(v)})$ a *polarized ideological state*. Hence, topic overlaps may lead to a dimensionality reduction in the underlying topic space: the opinion of an agent on one topic predicts their opinion towards another topic. With respect to our introductory example this means that a person who strongly opposes same-sex marriages will likely also be against transgender people using toilets of their identified gender.

The results shown in Figs.8.2(a)-(c) correspond to cases where the value of controversialness is equal for both considered topics. Relaxing this assumption provides us with additional flexibility. For instance, we may consider two topics where the first topic has a low value of α and the second topic is highly controversial. As we show in Fig. B.3(a) of App. B, this can result in the simultaneous emergence of consensus and opinion polarization with respect to the first and the second topic, respectively.

8.3 Mean-field approximation

To shed some more light on the states of consensus, uncorrelated polarization, and ideology we aim to capture the dynamics qualitatively within a mean-field approximation. This approach will allow us to understand the resulting dynamics in terms of the fixed points of a reduced

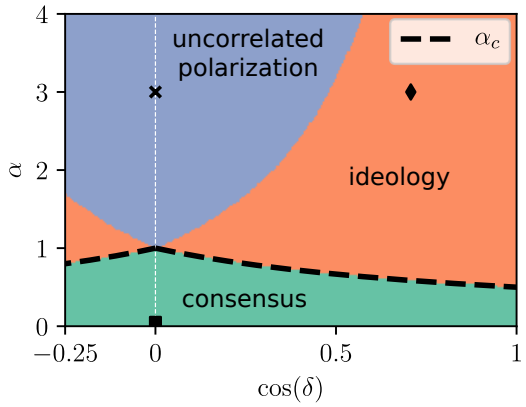

Figure 8.3: Phase diagram of opinion states obtained from the mean-field approximation, as a function of the topic overlap $\cos(\delta)$ and the controversialness α for $2Km\langle a\rangle = 1$. The colored phase space regions correspond to different emerging states: consensus (green), uncorrelated polarization (blue), and polarized ideological states (orange). The black dashed lines correspond to Eq. (8.11), i.e. the critical controversialness (α_c) separating the phases of consensus and polarization in two topic dimensions. Note that the phase diagram and α_c are symmetric with respect to the line of zero topic overlap, i.e. $\cos(\delta) = 0$ or $\delta = \pi/2$. The symbols are located at the parameter combinations of $\cos(\delta)$ and α used in Fig. 8.2 and Fig. 8.5.

model. To this end, we consider the thermodynamic limit of $N \to \infty$ and very high values of homophily $\beta \gg 1$. In these limits, we can assume that agents' opinions will be close to the ones of their interaction partners, i.e., we have $x_i^v \approx x_j^v \equiv x^v$. This approximation yields

$$\dot{x}^{(v)} = -x^{(v)} + 2Km\langle a\rangle \tanh\left(\alpha[\Phi\mathbf{x}]^{(v)}\right), \tag{8.5}$$

from Eq. (8.1). In the latter equation, we have replaced $A_{ij}(t)$ by the average number of interactions received by an agent in each time step, $2m\langle a\rangle$, which is a sum of two contributions: (i) the expected number of links per time step if the agent is activated ($\langle a\rangle m$), and (ii) the expected number of links an agent receives from all other activated agents, $\langle a\rangle \sum_{i\neq j} \frac{m}{N} = \langle a\rangle \frac{(N-1)m}{N}$, which is $\langle a\rangle m$ for large N. For fast-switching AD dynamics this approximation is well justified, as we have seen in Ch. 7 for $T = 1$. According to Eq. (8.5), the dynamics of the model is effectively described by a situation in which a single agent's opinion is driven by interactions with peers holding the same opinion, i.e., a self-interacting agent.

To derive a relation between the α and $\cos(\delta)$, which determines the transition between consensus and the emergence of opinion polarization, we proceed as in the one-dimensional case. First, we compute the Jacobian of Eq. (8.5) at $\mathbf{x} = 0$. For simplicity, we assume that all pairwise topic overlaps are equal, i.e., we have $\delta_{uv} = \delta\,, \forall u, v$. This is indeed a strong

assumption, but necessary to proceed analytically. The Jacobian of Eqs. (8.5) reads

$$\mathbb{J}(0) = \begin{pmatrix} -1+\Lambda\alpha & \Lambda\alpha\cos(\delta) & \dots & \Lambda\alpha\cos(\delta) \\ \Lambda\alpha\cos(\delta) & -1+\Lambda\alpha & \dots & \Lambda\alpha\cos(\delta) \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda\alpha\cos(\delta) & \Lambda\alpha\cos(\delta) & \dots & -1+\Lambda\alpha \end{pmatrix}, \tag{8.6}$$

where we have defined $\Lambda = 2Km\langle a\rangle$. The largest eigenvalue of $\mathbb{J}(0)$ is given by

$$\tilde{\lambda} = (T-1)(-1+2Km\langle a\rangle\alpha) + 2Km\langle a\rangle\alpha\cos(\delta), \tag{8.7}$$

and the full consensus ($\mathbf{x} = 0$) is stable for $\tilde{\lambda} < 0$. Solving the right hand side of Eq. (8.7) for α yields

$$\alpha_c = \frac{T-1}{2Km\langle a\rangle[T-1+\cos(\delta)]}, \tag{8.8}$$

relating the critical controversialness α_c to the overlap of topics, $\cos(\delta)$, for an arbitrary number of dimensions T.

For $T = 2$ setting Eqs. (8.5) to zero yields the following non-linear system of equations

$$\begin{aligned} x^{(u)} &= \tanh\left(\alpha[x^{(u)} + \cos(\delta)x^{(v)}]\right) \\ x^{(v)} &= \tanh\left(\alpha[\cos(\delta)x^{(u)} + x^{(v)}]\right). \end{aligned} \tag{8.9}$$

In panels (d)-(f) of Fig. 8.2, the corresponding deterministic attractors are depicted, where the values of α and $\cos(\delta)$ correspond to the ones used in the simulations of the full model, see panels (a)-(c). The resulting sets of fixed points are reminiscent of the steady states, obtained for the full stochastic model. For small controversialness (low α) there is only one fixed point, which corresponds to the global consensus at $\mathbf{x} = 0$. For increasing values of α the consensus is destabilized. In the case of orthogonal topics, four stable fixed points emerge that correspond to the uncorrelated polarized state [Fig. 8.2(e)]. For sufficiently large topic overlaps $\cos(\delta)$, the two fixed point with mixed stances, i.e. $[\mathrm{sgn}(x^{(u)}), \mathrm{sgn}(x^{(v)})] = \{(-,+),(+,-)\}$ are destabilized. As shown in panel (e), only the two fixed points with equal stances are stable, reflecting the formation of polarized ideological states in the full model.

These point-wise results can be extended to larger ranges in the α-$\cos(\delta)$ plane, giving rise to a phase diagram. In Fig. 8.3, we show the regions of stability for all three phases, i.e., consensus (green), polarization of uncorrelated opinions (blue), and ideology (orange). First, we note that the phase diagram is symmetric with respect to the line of vanishing overlaps ($\cos(\delta) = 0$), i.e., orthogonal topics, where no ideological states emerge, except for $\alpha = 1$: here the ideological phase (orange) extends until $\cos(\delta) = 0$. This gives rise to a triple point, where all three phases coincide, suggesting that ideological states may emerge even for infinitely small overlaps. This behavior can be understood in terms of the non-trivial fixed point solutions to Eqs. (8.5) for two topics ($T = 2$), u and v, and $2Km\langle a\rangle = 1$ at $\alpha = 1$. To find those solutions the $\tanh^{-1}(\cdot)$ is

taken on both sides of Eqs. (8.9) and the non-linearity is Taylor expanded up to third order, yielding

$$\begin{aligned}\frac{\left(x^{(u)}\right)^3}{3} &\simeq \cos(\delta)x^{(v)}\\ \frac{\left(x^{(v)}\right)^3}{3} &\simeq \cos(\delta)x^{(u)}.\end{aligned} \tag{8.10}$$

Equations (8.10) suggest that for positive overlaps $\cos(\delta) > 0$ the non-vanishing solutions $(x^{(u)*}, x^{(v)*} \ll 1)$ have equal signs, i.e., $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (+,+)$ or $(-,-)$. Specifically, close to the triple point the solutions behave like $(x^{(u)*}, x^{(v)*}) = \sqrt{3\cos(\delta)}(1,1)$, and $(x^{(u)*}, x^{(v)*}) = \sqrt{3\cos(\delta)}(-1,-1)$. Instead, in the case of negative overlaps, i.e., $\cos(\delta) < 0$ ($\delta \in]\pi/2, \pi[$), the stability of the system is reversed in the sense that the non-trivial solutions have opposite opinion stances: $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = (-,+)$ or $(+,-)$. Note however that this reversal, which is reflected in the symmetry of the phase diagram with respect to $\cos(\delta) = 0$, does not yield qualitatively new dynamics. It simply results in negatively correlated opinions, instead of positively correlated ones for $\cos(\delta) > 0$. This is demonstrated also for stochastic simulations of the full model in App. B. With regard to the analysis of empirical data in Sec. 8.6, negatively correlated opinions would simply arise from positively correlated ones by the reversal of the answer scale for one of the considered questions. In our discussion, we therefore omit negative topic overlaps and focus solely on cases with $\cos(\delta) > 0$.

From Fig. 8.3 we further note that increasing overlaps yield a wider region of stability for ideological states. The phase transition from uncorrelated polarized states to ideological states is critically driven by the topic overlap $\cos(\delta)$. The transition is sharp with respect to this parameter, such that by increasing the value of $\cos(\delta)$, the final configuration of the agents suddenly changes from uncorrelated polarization to the ideological phase. In App. B, we report details on the numerical procedure to generate the phase space diagram, shown in Fig. 8.3. Similarly, the transition between global consensus and ideology is driven by the controversialness α in a non-linear fashion. For $T = 2$, Eq. (8.8) reduces to

$$\alpha_c = \frac{1}{2Km\langle a\rangle[1+\cos(\delta)]}, \tag{8.11}$$

defining the critical controversialness α_c and the stability limits of the consensus phase for two topics. In Fig. 8.3, we depict Eq. (8.11) as black dashed line. Note that for orthogonal topics Eq. (8.11) reduces to Eq. (7.9) and therefore recovers the mean-field results of the one-dimensional model for $r = 1$. Importantly, the larger the overlap between topics, the smaller becomes the critical controversialness α_c necessary to de-stabilize consensus and promote ideological states.

8.4 Higher-dimensional case

Up to this point, the focus was on the simplest case with $T = 2$. In this section, we go beyond

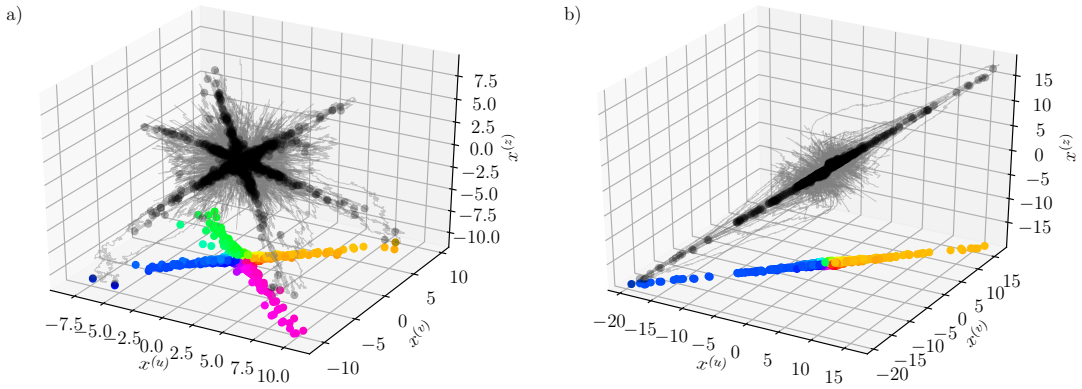

Figure 8.4: Temporal evolution of the opinions in a three-dimensional topic space ($T = 3$) for strong social influence ($K = 3$), controversial topics ($\alpha = 3$), and high homophily ($\beta = 3$). The grey lines and black dots correspond to the time evolution of agents' opinions and the steady states, respectively. In both cases, panels (a) and (b), topics u and v are orthogonal. In panel (a), all topics, including z, are pair-wise orthogonal, resulting in an uncorrelated polarized state, with weak correlations among the three opinions: $\rho(x^{(u)}, x^{(v)}) = 0.15$, $\rho(x^{(u)}, x^{(z)}) = 0.17$ and $\rho(x^{(v)}, x^{(z)}) = 0.11$. By contrast, in panel (b), topic z has a finite overlap with both topics u and v. Specifically, we have $\cos(\delta_{uz}) = \cos(\delta_{vz}) = \pi/4$. This leads to an ideological state, where the opinions with respect to the three topics (u, v, and z) are strongly correlated, and we find $\rho(x^{(u)}, x^{(v)}) \simeq \rho(x^{(u)}, x^{(z)}) \simeq \rho(x^{(v)}, x^{(z)}) \simeq 1$. For simplicity of illustrations the opinion space in panel (b) is depicted using orthogonal axes, despite the assumption of $\delta_{uz} = \delta_{vz} < \pi/2$.

the two-dimensional case and consider three topics u, v, and z. In the following, we investigate the interplay of the three involved topic overlaps, namely $\cos(\delta_{uv})$, $\cos(\delta_{vz})$, and $\cos(\delta_{uz})$. In particular, we are interested in the following question: if two topics u and v are orthogonal ($\cos(\delta_{uv}) = 0$), what is the effect of a third topic z with respect to the emergence of correlations between topics u and v?

Again we consider a regime of strong social influence ($K = 3$), high homophily ($\beta = 3$), and high values of controversialness ($\alpha = 3$). For $\cos(\delta_{vz}) = \cos(\delta_{uz}) = 0$, i.e. , if topic z neither overlaps with topic u nor with topic v an uncorrelated polarized state emerges, confirming the dynamical behavior observed in two dimensions. The three dimensional uncorrelated state is depicted in Fig. 8.4(a), where final opinions shown as black dots. The similarity to the two dimensional case, shown in Fig. 8.2(b), becomes apparent when we consider the projection of the three-dimensional opinion state on the two-dimensional (u, v)-plane: each dot corresponds to one agent's opinion, colored according to the opinion angle φ_i with respect to topics u and v. The projection reveals that opinion correlations with respect to topics u and v are very low, see caption of Fig. 8.4.

Now we consider a different situation, in which the third topic z has finite overlaps with the other two topics u and v, which, for simplicity, are assumed equal, i.e., we have $\cos(\delta_{uz}) = \cos(\delta_{vz}) > 0$. This results in a polarized ideological state, which is depicted in Fig. 8.4(b). It is

characterized by high opinion correlations with respect to topics u and v ($\rho_{uv} \simeq 1$), although they are orthogonal. Considering the state's projection to the (u, v)-plane, we find the opinions distributed precisely as in the case of the two-dimensional ideological state, cf. Fig. 8.2(c). This suggests that ideological states may emerge even with respect to entirely unrelated topics if the underlying topic space is expanded to higher dimensions and further related topics (here topic z) are introduced.

The results on the higher dimensional case have an interesting implication. They provide a possible explanation for the emergence of opinion correlations between two rather unrelated topics: those correlations might be due to the presence of further relevant topics, related to the previous two. While such confounders can be present, it may be hard to identify them in a given empirical data set. Our model suggests the search for such hidden topic dimensions in the case of high opinion correlations.

8.5 Opinion segregation on a network level

In this section, we investigate how the process of opinion formation is reflected in the co-evolving network of social interactions for $T = 2$. This discussion is somehow analogous to the one about the formation of echo chambers for $T = 1$ (Ch. 7). Independent of the topic dimension T, our model assumes that the opinion dynamics is coupled to the underlying network (via Eq. (8.3)), hence its structure is shaped by the dynamics described by Eq. (8.1). In Figs. 8.5(a), (b), and (c) we show interaction networks, which are obtained by the full model for the same set of parameters as used in panels (a)-(c) of Fig. 8.2. The depicted networks are established by the aggregation of 70 consecutive snapshots of the temporal adjacency matrix $A_{ij}(t)$, once a steady opinion state was reached. Each agent i corresponds to a node, whose size is proportional to the agent's conviction r_i. The color represents the opinion in the polar coordinate φ_i.

The network shown in Fig. 8.5(a) is obtained for a system approaching consensus. Although agents with similar opinions are likely to be connected – an effect, which is caused by homophily independent of the value of α – we do not observe a clear group structure forming between agents of different opinions (node colors). This drastically changes for opinion polarization. Panel (b) shows the network for the uncorrelated polarized state, where four clearly distinguishable opinion groups are visible, each one characterized by a different opinion (color code). Similarly, the ideological state manifests in a network structure, which is mainly segregated into two different opinion groups, see panel (c). We quantify these qualitative observations by a community detection analysis using the Louvain algorithm [173], which is based on the principle of modularity maximization, as discussed in Ch. 2. The bottom panels of Figs. 8.5(a), (b), and (c) show the community structure of the corresponding aggregated networks. Each community is represented by a polar bar, which is oriented according to the average opinion $\langle\varphi\rangle$ within the community. Their radii correspond to the community sizes and their colors and widths encode the average cosine similarity in a community, which is

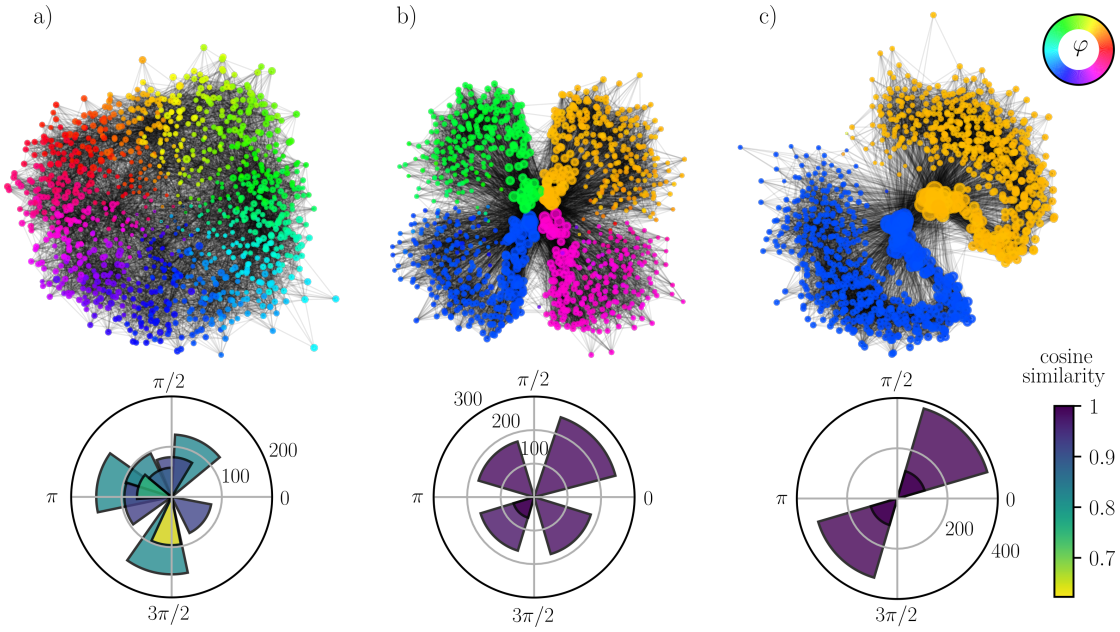

Figure 8.5: Correspondence between opinions and the structure of the aggregated social networks. The depicted networks (top panels) are aggregated over 70 time steps while the system approaches consensus (a), or is in a polarized steady state (b) and (c). The model parameters are set as in Fig. 8.2(a)-(c), i.e.: $\alpha = 0.05$, $\delta = \pi/2$ (a), $\alpha = 3$, $\delta = \pi/2$ (b) and $\alpha = 3$, $\delta = \pi/4$ (c). Each node corresponds to an agent, colored according to its opinion angle φ. The size of each node corresponds to the agent's conviction r. The community analysis of each network is shown as a polar bar plot in the corresponding bottom panel. Each community is represented by a bar with a radius equal to the size of the community. The color and the width of the bar correspond to the average cosine similarity between all pairs of agents in the respective community, and the average opinion angle in each community $\langle\varphi\rangle$ is encoded in the orientation of the bar. Communities containing less than 5% of the agents were disregarded.

defined as the mean scalar product between all pairs of opinions.

For the global consensus (panel (a)), many communities are identified, whose orientations are rather random. The communities are characterized by a heterogeneous distribution of opinions, which is reflected in the low values of the cosine similarity. The networks associated with polarized opinion states are strongly different. Here the average opinion within each community is aligned with the fixed points, shown in Figs. 8.2(e) and (f). For uncorrelated polarization, the communities are characterized by four typical average opinions, which directly relate to the four fixed points in Fig. 8.2(e). The opinions within each community are very homogeneous, as suggested by the high values of the cosine similarities. A similar behavior arises for the ideological phase, see panel (c). Here only two typical average opinions can be observed within large and very homogeneous communities.

8.6 Comparison to the 2016 American National Election Survey

Let us now contrast the model's behavior to empirical data. To this end, we investigate the degree of polarization and correlation between opinions with respect to different topic pairs based on data drawn from the 2016 American National Election Studies (ANES). Broadly, the main objective of the ANES studies, which have been conducted in the U.S. since 1946, is the analysis of public opinion and voting behavior in the U.S. presidential elections. More specifically, this is achieved by interviewing representative samples of U.S. citizens on a variety of topics, ranging from personal, to ethical, to political questions. The data provided by those surveys has been proven suitable for different empirical research efforts, including the characterization of long-term trends of opinion polarization [117, 309]. In our analysis, we included 67 questions with a total of 253,984 valid responses. See App. B for a more detailed discussion of the selection procedure of questions. Crucially, each respondent is assigned an ID such that the answers of a given respondent to different questions can be related to each other. We will analyse the ANES survey data with respect to two main features: (i) the (marginal) response distributions $P_u(x)$ for each considered question or topic u, and (ii) the absolute value of the Pearson correlation coefficient $\rho(u,v)$ between the responses to two questions or topics u and v. While the variance $\sigma_u^2(x)$ of $P_u(x)$ with respect to question u will be used in order to distinguish between consensus-like situations and highly polarized cases, $|\rho(u,v)|$ will reveal if the opinions with respect to two different issues u and v are correlated.

The subset of considered topics is illustrated in Fig. 8.6. The variances $\sigma_u^2(x)$ of the responses to each question u are plotted on top of panel (a). The questions are sorted in descending order of $\sigma_u^2(x)$, from highly polarized questions (left) to less polarizing ones (right). The majority of the response distributions show rather low values of $\sigma_u^2(x)$, a few, however, such as the question "voting is a duty" are strongly polarized. The main panel (a) depicts the correlation matrix of the responses (absolute values), sorted according to their variance $\sigma_u^2(x)$. The cell (u,v) is colored according to $|\rho(u,v)|$ of two questions u and v. While the average correlation value is approximately $\langle\rho\rangle = 0.2$, the full distribution of correlation values is rather broad

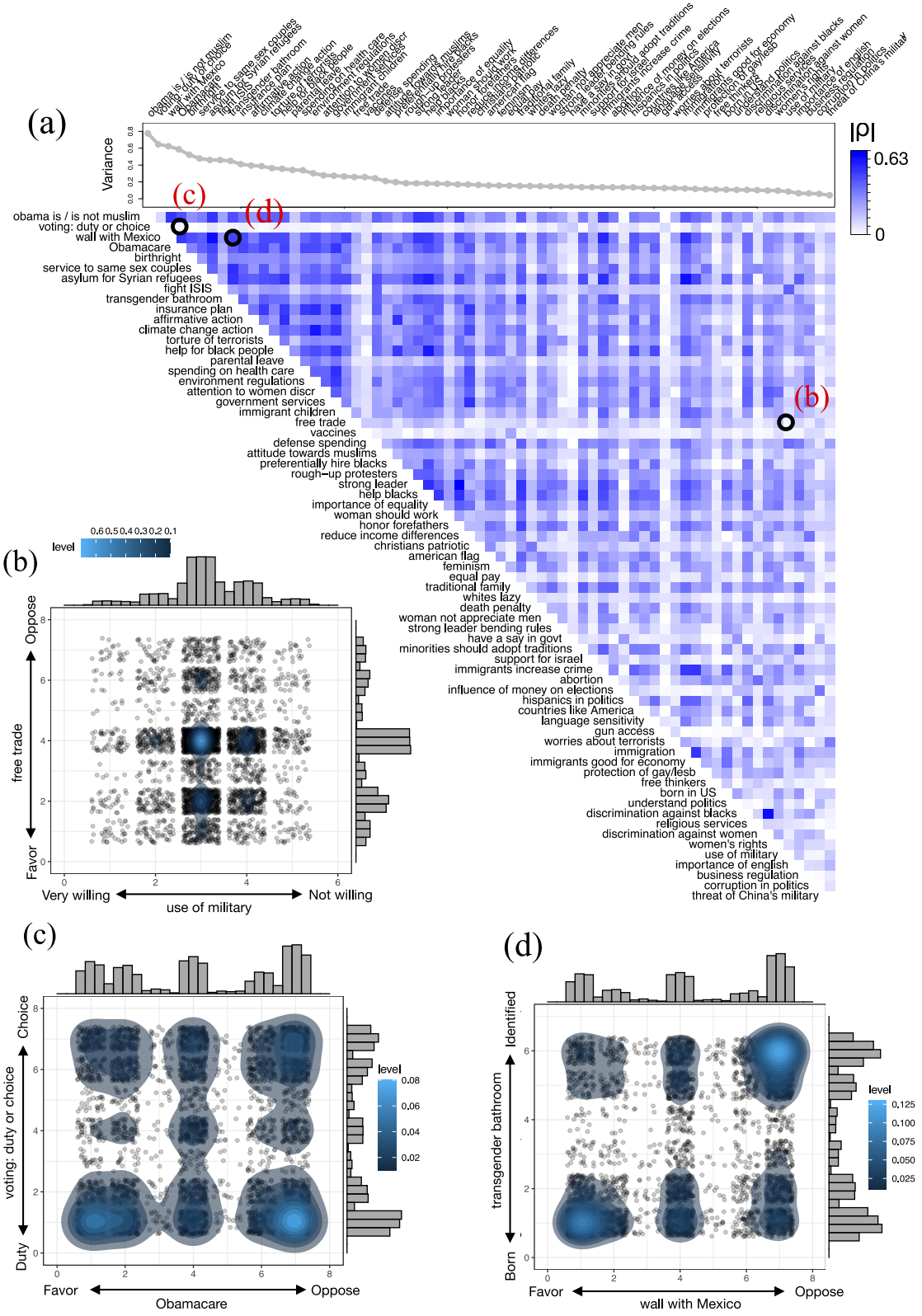

Figure 8.6: Analysis of responses to the ANES survey. Panel (a): Correlation matrix for all pairs of 67 analyzed questions from the ANES survey. Panels (a)-(c): Scatter plots of responses to different pairs of selected questions v and z: each dot represents one respondent by his/her responses to both questions.

with some pairs being strongly correlated, as reported in App. B. Although there is an overall dependence of the correlation values on the variance, low and high correlation values can be observed in all parts of the matrix. Thus, high correlation values are observed for question pairs with both small and high variances.

Three prototypical examples of two-dimensional opinion distributions, which correspond to the states observed for the model, are shown in the remaining panels (b)-(d): consensus-like (b), uncorrelated polarization (c), and polarized ideology (d). In Fig. B.6 of App. B, we show additional examples including the case where consensus is found with respect to one topic and opinion polarization occurs with respect to the other topic. This situation can be reproduced by the model if different values of α are considered for each topic, as discussed in Sec. 8.2 and App. B. The marginal response distributions $P_u(x)$ and $P_v(x)$ for the questions u and v are plotted on the x and y axis, respectively. For the sake of a clear visualization, the discrete data points are jittered, i.e. a small amount of noise is applied to each point to prevent overplotting. In panel (b), we show the consensus-like situation with low variances ($\sigma_u^2 = 0.08$, $\sigma_v^2 = 0.25$) and a low correlation ($\rho(u,v) = 0.02$) for the questions "Do you favor, oppose the U.S. making free trade agreements with other countries?" (answer on a 7 point-scale) vs. "How willing should the United States be to use Military force to solve international problems?" (5 point-scale). Both of these questions are peaked around a neutral opinion, and their response distributions exhibit a low variance. In panel (c), the responses to the questions "Do you consider voting a choice or a duty?" (7 point-scale) vs. "Do you favor, oppose the health care reform law passed in 2010?" (7 point-scale) (obamacare law), are both strongly polarized ($\sigma_u^2 = 0.58$, $\sigma_v^2 = 0.64$) but not correlated, corresponding to an uncorrelated polarized state ($\rho(u,v) = 0.03$). Finally, in panel (d), we show the case of two questions with polarized ($\sigma_u^2 = 0.62$, $\sigma_v^2 = 0.49$) and correlated responses ($\rho(u,v) = 0.44$), namely "Should transgender people have to use the bathrooms of the gender they were born as, or should they be allowed to use the bathrooms of their identified gender?" (6 point-scale) vs. "Do you favor, oppose building a wall on the U.S. border to Mexico?" (7 point-scale).

Strong opinion correlations may indeed be expected for closely related issues, such as the pair of questions regarding transgender people and same-sex couples. In Fig. B.6 of App. B, we show that the responses to these two questions are indeed highly correlated. By contrast, the high correlation value found for the question pair of "(...) building a wall on the U.S. border to Mexico" and transgender issues [Fig. 8.6(d)], comes as a surprise. In App. B, we show further examples of such less intuitive ideological states. Our model proposes a twofold explanation for opinion correlations with respect to rather unrelated topics. On the one hand, we may consider a low dimensional representation, where potential relations between two topics are encoded in the topic overlap $\cos(\delta)$. As we have shown in Sec. 8.3 by means of a mean-field analysis, ideological states may emerge for very small topic overlaps, driven by the controversialness of topics, homophilic interactions, and the assumed social reinforcement mechanism. On the other hand, additional topics may be considered in a higher dimensional topic space, which are not directly observed in the data. As shown in Sec. 8.4, such confounding topics may give rise to opinion correlations also for completely independent topic pairs.

8.7 Chapter summary and discussion

In this chapter, we generalized the one-dimensional radicalization model (Ch. 7) to multiple topic dimensions. In particular, this generalization implied that agents do not hold only one opinion towards a single topic, but T opinions towards T topics. Inspired by skewed coordinate systems, which were recently proposed in natural language processing applications, we assumed that opinions evolve in a multi-dimensional opinion space, which is spanned by the topics that are discussed simultaneously. While the mechanisms of social reinforcement and homophily are very similar to those assumed in the one-dimensional model, the multi-dimensional model contains a crucial extension: we assume that topics are not necessarily orthogonal. This assumption translates into pairwise topic overlaps that are stored in the matrix Φ. The formalism allows us to describe an empirical feature often observed on top of opinion polarization: issue alignment, where opinions are not only polarized, but show strong correlations. In our model, such polarized ideological states may emerge by relaxing the assumption of an orthogonal topic basis, i.e., if topics can thematically overlap. While such overlaps have a geometrical interpretation in our model, they may also be interpreted in terms of persuasive-arguments theory: certain arguments may hold for different topics, and the larger the set of such arguments, the larger is the overlap between topics. We analytically and numerically characterize the transitions between global consensus, uncorrelated polarized states, and polarized ideological states. Interestingly, we find within a mean-field approximation that ideologically aligned opinions may also arise between rather unrelated topics. This may either occur due to small overlaps for $T = 2$ or hidden confounding topics in higher dimensions, as shown for $T = 3$. The observed opinion states are in qualitative agreement with empirical data from the 2016 ANES. This finding suggests that our model may provide a further step towards the understanding of opinion correlations in polarized debates on controversial issues.

Previously, further multi-dimensional models were introduced that give rise to the phenomenon of issue alignment. In Ref. [310], the authors focus on the dynamics of work related opinions, which may determine the ability of work teams to perform certain tasks [311, 312]. The main ingredients of the model are homophily, social influence, rejection, and heterophobia, where the latter refers to the effect of two people disliking each other if their dissimilarity exceeds a certain threshold [313, 314]. More specifically, the model is defined as follows. Each team member is described by an agent, which is characterized by a set of binary and inflexible (demographic) attributes and another set of continuous and flexible attributes. Each of such flexible attributes corresponds to a work-related opinion of the agent. The social influence that two agents exert on each other is either assimilative or repulsive, if the distance in the combined space of the demographic attributes and the opinions is low or large, respectively. It is found that agents' opinions indeed polarize and align if the demographic attributes are highly correlated.

Social influence mechanisms based on assimilation and repulsion have been considered in a further multi-dimensional model, which gives rise to polarized and correlated opinion

distributions [114]. Specifically, the two-dimensional model assumes that each agent is characterized by a main opinion and a secondary one. The model has two important parameters: the rejection and the attraction threshold, which are defined in line with confidence bounds. If two agents meet and agree, i.e., their main opinions are within the agreement threshold, agents' opinions assimilate with respect to both dimensions. If, however, agents do not agree with respect to their main opinions and are, at the same time, too close to each other with respect to their secondary opinions (determined by the rejection threshold), those secondary opinions will reject each other and become more dissimilar. The model does not assume a specific network topology. Instead, in each time step two agents are picked randomly to simulate their encounter corresponding to a well-mixed population. For certain parameter values of the attraction and rejection thresholds the model gives rise to heterogeneous states with high opinion correlations.

A further recent attempt to explain the emergence of correlations in polarized opinion states is based on weighted balance theory (WBT) [296], which can be regarded as an extension to Heider's cognitive balance theory [35], see Ch. 1. The agent-based model derived from WBT generates opinion distributions which are polarized including correlated opinions towards different policy issues. Upon the random encounter of two agents the interaction and opinion update takes place as follows. First, an agent i determines their attitude towards agent j, which is denoted as $\mathscr{A}(i, j)$. Broadly, $\mathscr{A}(i, j)$ is derived from balance theory such that it reflects the overall cognitive consistency among two agents. Accordingly, the attitude is negative (positive) if most triads between both agents are evaluated negatively (positively) resulting in negative (positive) values of $\mathscr{A}(i, j)$. An important parameter of the model is the evaluative extremeness, which corresponds to how "black and white" the issue appears, similar to the controversialness α in Eq. (8.1). More specifically, for high evaluative extremeness agents dis(agree) more with other agents they in principle dis(agree) with. After evaluating the attitudes towards each other, agents adjust their opinions in order to maximize the cognitive balance with respect to their interpersonal attitudes. For sufficiently high values of evaluative extremeness the model gives rise to polarized and correlated opinion distributions.

In contrast to these previous approaches, our model does not assume a well-mixed population of agents, where contacts are sampled uniformly random across the population. Instead, interactions are ruled on homophily, where agents are more likely to interact with similar peers. Moreover, the mechanism of social influence in our model is different from assimilative opinion dynamics, rejection mechanisms, or the approach based on Heider's balance theory. We assume social reinforcement, where like-minded peers mutually increase their convictions, inspired by the law of group polarization and persuasive-arguments theory. Additionally, we assume that the time-evolution of opinions is connected to the dynamics of the temporal network of interactions. This co-evolution is induced by a homophily mechanism between the activity-driven network and the opinions dynamics. Except the different values of activity we disregard individual properties of agents, which is both a weakness and a strength of our approach. One the one hand, we do not take into account important features, such as demographic attributes, which may influence the overall opinion dynamics. On the other hand,

our model does not rely on those features to generate opinion correlations. Ideological states emerge due to quite general and realistic assumptions about social reinforcement, homophily and partially overlapping topics. Such overlaps are not a purely theoretical construct. Future research should therefore further investigate the connection between two (independent) empirical observations: (i) the correlation between opinions with respect to different topics (quantified by surveys or extracted from online social media), and (ii) the thematic overlap between these two topics. Quantifying such overlaps could be addressed by topic modeling of large data sets related to the topics under consideration [315].

9 Conclusions and outlook

At least for the past 70 years, since Robert Abelson formulated his paradox on the formation of opinion cleavages, persistent social division has puzzled researches ranging from sociologists, to social psychologist, to physicists. What are the mechanisms behind the emergence of population-scale consensus and agreement, or persisting opinion polarization? In this thesis, we have tackled this question using mathematical models to bridge the micro-macro gap between empirically motivated mechanisms of social influence and the collective opinion states observed in large populations of interacting individuals. The presented models can be broadly divided into two classes, which are reflected in the structure of the thesis.

In Part I, we focused on assimilative consensus models, which have previously been applied both in the context of opinion formation and in technical applications such as multi-vehicle formation control. Generally, assimilation models imply that the (opinion) states of agents become increasingly similar upon interactions. On connected networks, it follows that consensus is inevitably reached and fostered the *more* – and the *more intensely* – agents (socially) interact and influence each other. Using two approaches, based on the spectral decomposition of the system's network Laplacian, we examined mechanisms which inhibit the formation of consensus states although the network of interactions was connected.

In Ch. 4, we investigated the effects of stubborn agents in an Abelson-type assimilation model. In contrast to regular agents, stubborn agents adhere to specific opinion values and therefore have a significant impact on the collective opinion dynamics. In particular, we focused on (i) the change of a prevailing consensus due to a single stubborn agent, and (ii) the emergence of heterogeneous opinions in the case of a pair of antagonistically biased agents. The spectral approach allowed us to describe the assimilative opinion dynamics and the underlying social influence network in a unified framework. Notably, in the case of two antagonistic stubborn agents the emerging heterogeneous opinion states could be described in terms of a modified version of the resistance distance, which uncovered links between the resulting opinion formation processes and structural properties of the social influence networks. The findings shed light on the interplay between network topology and the placement of influential actors, with potential implications for influence maximization in social networks.

In Ch. 5, consensus formation was studied outside of opinion dynamics. Specifically, we considered second-order consensus models with time-periodic coupling modulations. Second-order consensus models take into account the Newtonian dynamics of its constituent agents and have previously been applied in technical applications, for instance, as decentralized protocols to achieve consensus in multi-vehicle systems. For certain intermediate coupling frequencies we uncovered a mechanism, based on parametric resonance, which inhibits the formation of consensus and promotes an exponential excitation of the system. This behavior is at odds with the expected emergence of consensus for very fast and slow coupling time scales. The spectral decomposition approach allowed us to predict the (parametric) resonance frequencies and to link them to the Laplacian spectrum of the underlying (static) backbone network. Our results may have implications for real-world consensus applications, where the coupling functions can be subject to external and periodic perturbations.

In Part II, we departed from classical models of assimilation and presented models of opinion dynamics, in which frequent interactions and high social influence strengths do *not* necessarily lead to the faster emergence of consensus, nor its increased stability. Quite the contrary: too many interactions, or too strong social influence among agents, may indeed destabilize consensus states. Inspired by findings on echo chambers in online social networks, we proposed a minimal model based on a non-linear social reinforcement mechanism. It gives rise to a radicalization dynamics, where opinions may become more extreme starting from moderate initial conditions. Within this minimal approach, we were able to reproduce qualitatively certain stylized facts found in empirical data on polarized political debates without relying on the previously criticized assumption of opinion differentiation.

In Ch. 7, the model was introduced in one dimension, where agents hold an opinion on a single topic. We implemented a minimal mechanism for group polarization, where agents' convictions become more extreme upon social interactions with like-minded agents. In addition to social reinforcement, we considered interactions ruled by homophily and agents, which are characterized by different levels of activity. Microscopically, the model mimics persuasive-arguments theory, where two interacting agents holding equal opinion stances reinforce their opinions, while their opinions converge towards the neutral stance for opposing opinions. On a macro-scale, the model gives rise to qualitatively different opinion states: a global consensus for rather uncontroversial topics and low social interaction strengths, and (one-sided) radicalized states for low homophily, controversial topics, and strong social interactions. In the regime of high homophily, where a bi-polarized opinion state emerges for controversial topics, we contrasted the model to empirical data from Twitter. In addition to a pronounced bimodal opinion distribution, the model qualitatively captures (i) a clear association between users' activities and their opinions, and (ii) the opinion segregation on a network level, i.e., echo chambers. By investigating the transitions between the three states, we identify potential factors, which may destabilize consensus in discussions on social media, drive radicalization phenomena, and lead to opinion polarization.

In Ch. 8, we paid particular attention to the emergence of issue alignment. Although it can

be observed in many polarized debates, where extreme opinions on different topics often show significant correlations, issue alignment has received only little attention in previous modeling studies. How can such correlations emerge, without assuming them a priori in the individual preferences or in a preexisting social structure? Based on the assumption that topics are generally not discussed in isolation, we approach this question by generalizing the one-dimensional radicalization model (Ch. 7) to multiple dimensions. Specifically, we assumed that the opinion evolution unfolds in a multi-dimensional space, whose basis vectors correspond to the topics discussed and these topics are not necessarily orthogonal. In this picture, the angles between two base vectors translate into topical overlaps. Such overlaps have a geometrical interpretation with respect to persuasive-arguments theory in multi-dimensional topic spaces: the larger the geometric overlap between two topics, the higher is the probability that an argument supporting one topic also supports the other one. In the regime of large homophily, the model reproduces qualitatively two-dimensional opinion distributions as measured in the 2016 American National Election Studies (ANES), including consensus and opinion polarization. Most importantly, we found that by relaxing the assumption of an orthogonal basis of the topic space, i.e. if topics can overlap, polarized ideological states may emerge. Similar to states found in the ANES data, the model thus gives rise to states where the opinions with respect to different topics show significant correlations. Crucially, within a mean-field analysis we demonstrated that ideological states may also be formed for very small overlaps, driven by the interplay of social reinforcement and homophily. These insights may help to understand the observed opinion correlations with respect to rather unrelated but highly controversial topics.

Wrapping it up

We have demonstrated for different agent-based models how the system-wide outcomes in social systems depend on the network of interactions and the details of social influence. Therefore, it is our hope that this work can contribute to the understanding of the emergence of collective opinion states in order to shed further light on the micro-macro gap of social systems.

Outside of opinion dynamics, we investigated second-order consensus models with time-periodic coupling modulations and established a novel parametric resonance phenomenon on networks. Most importantly, however, our goal was to contribute to the understanding of collective opinion formation processes. Besides studying the effects of stubborn agents in a highly theoretical model, we aimed at understanding real-world phenomena based on empirical data. To this end, we developed a minimal and interpretable model of opinion dynamics to reproduce some observed stylized facts of empirical data drawn from Twitter and the 2016 American National Election Studies. In particular, we could show that the interplay of social reinforcement and homphily can give rise to several empirical features of opinion distributions, such as the existence of extreme opinions, bi-polarization, a pronounced association between the activities and the opinions of users, echo chambers, and ideological states.

Below, in the final section, we take a brief look forward and speculate on how our results may relate to some future developments in the field of computational social science.

What there might be next

Our "life in the network" [60] is characterized by increasing connectivities and ramped up information flows between individuals. These developments gave rise to new modes of public opinion formation. Social media have played a crucial role in this process. Approximately 60% of the three billion internet users utilize such platforms to communicate and as major sources of information [316, 317]. Despite potential positive aspects [13], social media have recently been associated with negative social developments. Although networks - by definition - connect their entities, this does not seem to be entirely true for platforms like Facebook or Twitter. This is reflected in the general view that social cohesion has not only benefited from social media. Instead, such platforms may accelerate detrimental societal developments, such as political radicalization and opinion polarization, and thereby contribute to the increase of social cleavages. Shortly after the shutdown of the social networking site Parler, it became clear to the public that the storming of the U.S. Capitol in January 2021 was not only well documented on the platform, but that it probably contributed substantially to the event. On Parler and other social media platforms used by the far-right, conspiracy theories and radicalizing content, including calls for violence on the members of the congress, have been circulating already months before the incident [318].

Not just since these recent events on Capitol Hill in Washington D.C., there is an ongoing debate about how to mitigate the negative effects of social media without restricting people's fundamental rights, such as the freedom of speech. For instance, while it is not forbidden to disseminate information to a large audience in order to manipulate public opinion formation, we should leave no stone unturned to understand such processes and mitigate their negative consequences on society. Recently, potential solutions were proposed within a call for behavioral interventions in online environments [319]. More specifically, it has been argued that interventions such as "nudges" [306] or "boosts" [320] may be used on social media. By either altering the choice architecture of individuals or by improving their competences with respect to certain tasks such interventions may help to promote truth and democratic discourse online, while preserving people's autonomy [319]. For instance, it is suggested that social media sites could provide better cues to their users such that they are enabled to distinguish easier between false and correct information. Another proposed option was to introduce some harmless but often effective psychological friction to decelerate the cycles of content production and consumption. Asking people if they have read the article they want to share, or checked the information contained in a post, does certainly not constitute a restriction of their fundamental rights, but it could potentially help to slow down the spread of misinformation or highly controversial content.

To implement such interventions reasonably and effectively it is important to better under-

stand the behavior of people online, their interactions, and the resulting effects on public opinion formation. Computational social science can make further major contributions to this effort. On the one hand, agent-based models, such as the ones presented in this thesis, may not only contribute to a better understanding of increasingly self-organized processes of opinion formation. They also yield flexible and cheap means to test interventions and investigate their macroscopic effects. For instance, a recent preprint extended the radicalization model, which we proposed in Ch. 7 and 8, by an additional term to model a nudging intervention [305]. The introduction of the term could reduce the system's bi-polarization and mitigated the strength of echo chambers. This example illustrates how, in principle, agent-based models can be used to test possible interventions in online environments. Insights gained from such simple models can, however, only be a first step.

As a final perspective, we therefore briefly discuss possibilities to overcome some limitations regarding the predictive power of agent-based models of social systems. On the one hand, this may be achieved by more extensive social psychological experiments. With regard to opinion dynamics, it is important to experimentally validate crucial assumptions about social influence. Previously, it was pointed out that details on even basic model ingredients such as confidence bounds lack profound empirical testing [3]. On the other hand, social science research may also step further out of laboratories and investigate human behaviors in realistic settings, i.e., directly on social media platforms. A recent study, for example, reported evidence of a complex contagion phenomenon that was found using bots on Twitter, providing insights on the content-sharing behavior of real users on social media [277]. Similar approaches are conceivable in order to investigate more complex processes, including the information exchange on political issues and the induced time-resolved opinion changes. Very recently, the potential of integrating two established paradigms in social science research has been pointed out [276]. It was argued that it is beneficial to combine (i) more traditional survey based research, and (ii) the analysis of digital trace data, as commonly performed in computational social science. The advantages of this combined approach are revealed by some limitations of our work. In particular, the "opinions" of Twitter users considered in Ch. 7 on controversial topics were inferred based on links to news websites with a known political leaning. While this is of course a rather crude approximation of an individual's real opinion towards a certain topic, the Twitter data contains information about the social interactions of users. By contrast, the reponses to specific questions in the ANES data set, analyzed in Ch. 8, are much more reliable. However, the ANES data lacks any information about the social network of respondents. Reliable opinion data combined with information on the social interactions of users bears great potential for the investigation of public opinion formation. Such studies could either be performed on existing social media platforms, where users are additionally and individually asked about their opinions on certain topics, or on platforms simulating social media for research purposes, such as the Truman platform developed at Cornell University [321]. Especially, promising seems the possibility to gather time-resolved opinion data, by which one could not only compare static opinion distributions with steady states of a model, but also try to match the "real" dynamics towards consensus or polarization.

A Appendix Part I

Derivation of relation 1

Elementwise, the product of $\mathbb{L}^{(\kappa)}$ and its inverse reads

$$\sum_j \mathbb{L}^{(\kappa)}_{ij} [\mathbb{L}^{(\kappa)}]^{-1}_{jk} = \sum_j \left(\mathbb{L}_{ij} + \kappa \sum_{a \in V_s} \delta_{ai} \delta_{ij} \right) [\mathbb{L}^{(\kappa)}]^{-1}_{jk} \tag{A.1}$$

$$= \delta_{ik} \,. \tag{A.2}$$

Summing over i, we get

$$\sum_{a \in V_s} [\mathbb{L}^{(\kappa)}]^{-1}_{ja} = 1/\kappa \,. \tag{A.3}$$

The matrices $\mathbb{L}^{(\kappa)}$ and $[\mathbb{L}^{(\kappa)}]^{-1}$ can be written using the eigenvectors and eigenvalues of $\mathbb{L}^{(\kappa)}$ as

$$\mathbb{L}^{(\kappa)}_{ij} = \sum_\alpha \lambda^{(\kappa)}_\alpha u^{(\kappa)}_{\alpha,i} u^{(\kappa)}_{\alpha,j} \,, \tag{A.4}$$

$$\left[\mathbb{L}^{(\kappa)}\right]^{-1}_{ij} = \sum_\alpha \frac{u^{(\kappa)}_{\alpha,i} u^{(\kappa)}_{\alpha,j}}{\lambda^{(\kappa)}_\alpha} \,. \tag{A.5}$$

It follows that,

$$\sum_{a \in V_s} \sum_\alpha \frac{u^{(\kappa)}_{\alpha,k} u^{(\kappa)}_{\alpha,a}}{\lambda^{(\kappa)}_\alpha} = 1/\kappa \,. \tag{A.6}$$

Derivation of relation 2

In the case of two stubborn agents $V_s = \{a, b\}$, we have for the element (a, b) of the product of $\mathbb{L}^{(\kappa)}$ by its inverse,

$$\sum_j \mathbb{L}^{(\kappa)}_{aj} [\mathbb{L}^{(\kappa)}]^{-1}_{jb} = \sum_j \mathbb{L}_{aj} [\mathbb{L}^{(\kappa)}]^{-1}_{jb} + \kappa [\mathbb{L}^{(\kappa)}]^{-1}_{ab} = 0 \,. \tag{A.7}$$

Exploiting the symmetry of matrices $\mathbb{L}^{(\kappa)^{-1}}$ and $\mathbb{L}$ together with Eq. (A.3) we get,

$$[\mathbb{L}^{(\kappa)}]^{-1}_{aa} = [\mathbb{L}^{(\kappa)}]^{-1}_{bb}\,, \tag{A.8}$$

$$\sum_j \mathbb{L}_{aj}[\mathbb{L}^{(\kappa)}]^{-1}_{jb} = \sum_j \mathbb{L}_{bj}[\mathbb{L}^{(\kappa)}]^{-1}_{ja}\,. \tag{A.9}$$

With Eq. (A.4) we have the relation between eigenvectors and eigenvalues,

$$\sum_\alpha \frac{{u^{(\kappa)}_{\alpha,a}}^2}{\lambda^{(\kappa)}_\alpha} = \sum_\alpha \frac{{u^{(\kappa)}_{\alpha,b}}^2}{\lambda^{(\kappa)}_\alpha}\,. \tag{A.10}$$

Derivation of relation 3

Multiplying $\mathbb{L}^{(\kappa)}$ by eigenvector $\mathbf{u}^{(\kappa)}_\alpha$ reads elementwise

$$\sum_j \mathbb{L}^{(\kappa)}_{ij} u^{(\kappa)}_{\alpha,j} = \sum_j \left(\mathbb{L}_{ij} + \delta_{ij}\kappa \sum_{a\in V_s}\delta_{aj}\right) u^{(\kappa)}_{\alpha,j} \tag{A.11}$$

$$= \lambda^{(\kappa)}_\alpha u^{(\kappa)}_{\alpha,i}\,. \tag{A.12}$$

Summing over i we get

$$\kappa \sum_j \sum_{a\in V_s}\delta_{aj}\, u^{(\kappa)}_{\alpha,j} = \lambda^{(\kappa)}_\alpha \sum_i u^{(\kappa)}_{\alpha,i} \tag{A.13}$$

which finally yields

$$\kappa \sum_{a\in V_s} \frac{u^{(\kappa)}_{\alpha,a}}{\lambda^{(\kappa)}_\alpha} = \sum_i u^{(\kappa)}_{\alpha,i}\,. \tag{A.14}$$

Positive eigenvalues of the modified Laplacian

The positivity of the eigenvalues of $\mathbb{L}^{(\kappa)}$ follows from

$$\begin{aligned}\mathbf{u}^{(\kappa)T}_\alpha \mathbb{L}^{(\kappa)}\mathbf{u}^{(\kappa)}_\alpha &= \sum_{i,j} u^{(\kappa)}_{\alpha,i}\left(\mathbb{L}_{ij} + \kappa\sum_{a\in V_s}\delta_{ai}\delta_{ij}\right)u^{(\kappa)}_{\alpha,i}\\ &= \mathbf{u}^{(\kappa)}_\alpha \mathbb{L}\mathbf{u}^{(\kappa)}_\alpha + \kappa\sum_{a\in V_s}{u^{(\kappa)}_{\alpha,a}}^2\,,\\ &= \lambda^{(\kappa)}_\alpha > 0\,,\end{aligned} \tag{A.15}$$

where in the last line we used that the Laplacian $\mathbb{L}$ is positive semidefinite.

Derivation of the opinion association relation

The final state of opinions $\mathbf{x}(t \to \infty)$ are obtained by a spectral decomposition over the eigenvectors of $\mathbb{L}^{(\kappa)}$. In the case of two antagonistic stubborn agents, denoted as a, b with $P_a = -P_b = P$, and vanishing initial conditions the expansion coefficients are given by

$$c_\alpha(t \to \infty) = \kappa P \frac{(u^{(\kappa)}_{\alpha,a} - u^{(\kappa)}_{\alpha,b})}{\lambda^{(\kappa)}_\alpha}, \quad \alpha = 1, ..., n. \tag{A.16}$$

Accordingly, the final opinion of agent i reads,

$$x^\infty_i = \sum_\alpha c_\alpha(t \to \infty) u^{(\kappa)}_{\alpha,i} \tag{A.17}$$

$$= \kappa P \sum_\alpha \frac{(u^{(\kappa)}_{\alpha,a} - u^{(\kappa)}_{\alpha,b})}{\lambda^{(\kappa)}_\alpha} u^{(\kappa)}_{\alpha,i}. \tag{A.18}$$

Using the definition of the MRD, the latter equation can be expressed as

$$x^\infty_i = \frac{\kappa P}{2} \sum_\alpha \frac{(2u^{(\kappa)}_{\alpha,a} u^{(\kappa)}_{\alpha,i} - 2u^{(\kappa)}_{\alpha,b} u^{(\kappa)}_{\alpha,i})}{\lambda^{(\kappa)}_\alpha} \tag{A.19}$$

$$= \frac{\kappa P}{2} [\Omega^{(\kappa,1)}_{bi}(\{a,b\}) - \Omega^{(\kappa,1)}_{ai}(\{a,b\})], \tag{A.20}$$

where we used that $\sum_\alpha \frac{{u^{(\kappa)}_{\alpha,a}}^2}{\lambda^{(\kappa)}_\alpha} = \sum_\alpha \frac{{u^{(\kappa)}_{\alpha,b}}^2}{\lambda^{(\kappa)}_\alpha}$, as derived in App. A.

Derivation of the limit $\kappa \to \infty$

In the case of two antagonistic stubborn agents with finite stubbornness κ, the opinion distance between the pair of stubborn agents satisfies $D_{\max} \leq |P_a - P_b|$. For $\kappa \to \infty$, each stubborn agent reaches their own final bias and thus the distance between them should satisfy $D_{\max} = |P_a - P_b|$. The latter can be shown as follows. From Eqs. (A.6), (A.14) one has

$$\sum_\alpha \frac{{u^{(\kappa)}_{\alpha,a}}^2 + u^{(\kappa)}_{\alpha,a} u^{(\kappa)}_{\alpha,b}}{\lambda^{(\kappa)}_\alpha} = 1/\kappa, \tag{A.21}$$

$$\sum_\alpha \frac{{u^{(\kappa)}_{\alpha,b}}^2 + u^{(\kappa)}_{\alpha,a} u^{(\kappa)}_{\alpha,b}}{\lambda^{(\kappa)}_\alpha} = 1/\kappa, \tag{A.22}$$

$$\frac{u^{(\kappa)}_{\alpha,a} + u^{(\kappa)}_{\alpha,b}}{\lambda^{(\kappa)}_\alpha} = \kappa^{-1} \sum_i u^{(\kappa)}_{\alpha,i}. \tag{A.23}$$

Then using the definition of the MRD, Eq. (4.10), with $p = 1$ one has,

$$\begin{aligned} D_{\max} &= \kappa P \Omega_{ab}^{(\kappa,1)} && \text{(A.24)} \\ &= \kappa P \sum_{\alpha} \frac{(u_{\alpha,a}^{(\kappa)} - u_{\alpha,b}^{(\kappa)})^2}{\lambda_{\alpha}^{(\kappa)}} && \text{(A.25)} \\ &= \kappa P \sum_{\alpha} \frac{{u_{\alpha,a}^{(\kappa)}}^2 - 2u_{\alpha,a}^{(\kappa)} u_{\alpha,b}^{(\kappa)} + {u_{\alpha,b}^{(\kappa)}}^2}{\lambda_{\alpha}^{(\kappa)}} && \text{(A.26)} \\ &= \kappa P \left[2 \sum_{\alpha} \frac{{u_{\alpha,a}^{(\kappa)}}^2 + {u_{\alpha,b}^{(\kappa)}}^2}{\lambda_{\alpha}^{(\kappa)}} - \frac{2}{\kappa} \right] . && \text{(A.27)} \end{aligned}$$

Finally, taking stubbornness κ to infinity one has

$$\lim_{\kappa \to \infty} D_{\max} = 2P , \qquad \text{(A.28)}$$

where we have used $\lim_{\kappa\to\infty} \sum_{\alpha} \frac{{u_{\alpha,a}^{(\kappa)}}^2}{\lambda_{\alpha}^{(\kappa)}} = 1/\kappa$.

Derivation of μ_x and σ_x^2

To obtain the final mean opinion in the case of two opposed stubborn agents, we average Eq. (4.20) over all agents' opinions. This yields

$$\begin{aligned} \mu_x(\{a,b\}) &= N^{-1} \sum_i x_i^{\infty} \\ &= N^{-1} \sum_i \frac{\kappa P}{2} [\Omega_{bi}^{(\kappa,1)}(\{a,b\}) - \Omega_{ai}^{(\kappa,1)}(\{a,b\})] && \text{(A.29)} \\ &= \frac{\kappa P}{2} [C_1^{-1}(b) - C_1^{-1}(a)] , \end{aligned}$$

In a similar manner, we can derive an expression for the variance of the final opinion distribution. It is given by

$$\begin{aligned} \sigma_x^2(\{a,b\}) &= N^{-1} \sum_i (x_i^{\infty} - \mu_x)^2 && \text{(A.30)} \\ &= \left(\frac{\kappa P}{2} \right)^2 \sum_i \left[\Omega_{ib}^{(\kappa,1)} - \Omega_{ia}^{(\kappa,1)} - C_1^{-1}(b) + C_1^{-1}(a) \right]^2 . && \text{(A.31)} \end{aligned}$$

To obtain Eq. (4.23), we note that

$$\begin{aligned} \frac{1}{4} \sum_i \left(\Omega_{ib}^{(\kappa,1)} - \Omega_{ia}^{(\kappa,1)} \right)^2 &= \sum_{i,\alpha} \frac{(-u_{\alpha,i}^{(\kappa)} u_{\alpha,b}^{(\kappa)} + u_{\alpha,i}^{(\kappa)} u_{\alpha,a}^{(\kappa)})^2}{{\lambda_{\alpha}^{(\kappa)}}^2} && \text{(A.32)} \\ &= \Omega_{ab}^{(\kappa,2)} . && \text{(A.33)} \end{aligned}$$

where we have used $\sum_\alpha \frac{u_{a,a}^{(\kappa)\,2}}{\lambda_\alpha^{(\kappa)}} = \sum_\alpha \frac{u_{a,b}^{(\kappa)\,2}}{\lambda_\alpha^{(\kappa)}}$.

Stochastic block model and friendship network

Stochastic block model (SBM): For the simple case considered in Ch. 4 we assume SBM networks with two blocks ($b = 2$) of equal size ($N/2$). In order to tune the community structure of the network, while fixing the average number of edges in the network, we implement the following relation

$$p_{\text{inter}} = \frac{N_E - [(N/2)^2 - N/2]\, p_{\text{intra}}}{(N/2)^2}\,. \tag{A.34}$$

Note that, for the considered case of $N = 50$ and $N_E = 200$, the SBM network disintegrates in two parts in the limit of $p_{\text{intra}} \to 1/3$, as $p_{\text{inter}} \to 0$.

Empirical friendship network. The considered data set contains information about friendships within a US highschool [244]. The friendship network was constructed by asking each student twice about their friends in the same school. Accordingly, the original network is directed and weighted to account for multiple namings of a single student by a friend. Here, we symmetrize the network $\mathbf{W}$ and dismiss weights such that we have $W_{ij} = 1$ if one of the two students (i, j) named the other as a friend and $W_{ij} = 0$, otherwise. The network contains $N = 70$ nodes and $N_E = 274$ edges. To randomize the network topology we perform an increasing number of double edge swaps. Importantly, this procedure fixes the degrees of the nodes but randomizes the connectivity structure [322].

Opinion coherence and complexity of numerical algorithms

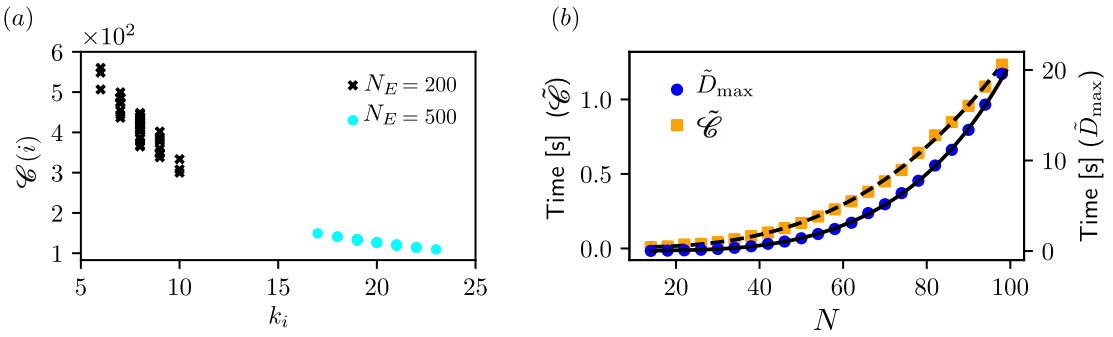

Figure A.1: Panel (a) shows the relation between the coherence measure $\mathscr{C}(i)$ and the degree k_i of the stubborn agent's node for two different WS networks with $N = 50$ nodes and $N_E = 200$ (black crosses) and $N_E = 500$ (cyan dots) and a rewiring probability of $p_{\text{ws}} = 0.2$ in both cases. In panel (b) we depict the time complexity in seconds [s] for the computation of the quantities $\tilde{\mathscr{C}}$ (orange squares) and $\tilde{D}_{\max}$ (blue dots) as a function of the network size N. The black dashed and solid lines correspond to polynomial fits of the orders $\mathscr{O}(n^3)$ ($\tilde{\mathscr{C}}$) and $\mathscr{O}(n^5)$ ($\tilde{D}_{\max}$), respectively.

In panel (a) of Fig. A.1, the relation between the opinion coherence $\mathscr{C}(i)$ and the degree of node i, k_i, on which the stubborn agent is placed is depicted. The black crossed markers show the results, which are also depicted in the top panel of Fig. 4.2(b). By contrast, the cyan dots show results of a WS network with an increased number of $N_E = 500$ edges, where the resulting coherence measures $\mathscr{C}(i)$ overlap less for different values of k_i.

In Fig. A.1(b), we explore the time complexity for the numerical computation of $\tilde{\mathscr{C}}$ and $\tilde{D}_{\max}$, which are computed as averages over all possible sets of stubborn agents on a WS networks with N nodes and $K_{\mathrm{WS}} = 4$. For simplicity we set $p_{\mathrm{ws}} = 0$. Due to the quadratic growth of the possible sets of stubborn agents as a function of N the time complexity of $\tilde{D}_{\max}$ (blue dots) is increased about a factor of N^2 compared to $\tilde{\mathscr{C}}$ (orange squares).

B Appendix Part II

Numerical simulations of the radicalization model

For the numerical simulation of the radicalization model (in one or more topic dimensions T) we fix the number of agents N, controversialness α, homophily β, social interaction strength K and the AD params (m, ε and γ). In multiple dimensions ($T > 1$) it is also necessary to specify the pairwise angles between all considered topics u and v, δ_{uv}, i.e., the topic overlap matrix Φ. The activities of individual agents are drawn from the distribution $F(a) = [(1-\gamma)/(1-\varepsilon^{1-\gamma})]a^{-\gamma}$ and are fixed for each simulation run. The initialization of opinion states is specified in Ch. 7 and 8. The temporal network $A_{ij}(t)$ and the agents' opinions are updated in each time step as follows:

(i) First, in each time step, the network is represented by N disconnected nodes. Accordingly, $A_{ij}(t)$ is the zero matrix. Then each agent is activated with probability a_i.

(ii) If active, agent i contacts m distinct agents. Importantly, the probability that agent i contacts agent j is given by p_{ij}, which is defined in Eq. (7.3) and Eq. (8.3) for the one- and multi-dimensional model, respectively. This contact is represented by the directed edge (j,i) in the temporal adjacency matrix, i.e., $A_{ji}(t) = 1$. With probability r (reciprocity) the link (j,i) is reciprocal, such that there is also social influence from agent j to agent i, i.e., $A_{ij} = 1$. Note that agent i chooses m agents based on p_{ij} without replacement, such that agent i cannot contact agent j multiple times in a single time step. In the multidimensional model (Ch. 8) we assume $r = 1$, and thus each established link is reciprocated.

(iii) After the temporal matrix $A_{ij}(t)$ is formed, each agent receives their aggregated social input coming from their neighbors, and the opinions are updated by numerically integrating Eq. (7.1) (for $T = 1$, cf. Ch. 7) or Eq. (8.1) (for $T > 1$, cf. Ch. 8) using a Runge-Kutta fourth-order method [323]. After the opinion updates the process starts anew from step (i).

Additional results for the one-dimensional model

In Fig. B.1 we depict results complementary to those presented in Ch. 7 for $r = [0.1, 0.3, 0.65, 0.9]$. The results are robust with respect to variations of the reciprocity r. For both large and small values of r opinion bi-polarization occurs [panel (a)], echo chambers emerge [top row of panel (c)], and there is a clear association between the agents' activities and their opinions [bottom row of panel (c)]. Furthermore, the mean-field approximation captures well the transition from consensus to radicalization [panel (b)].

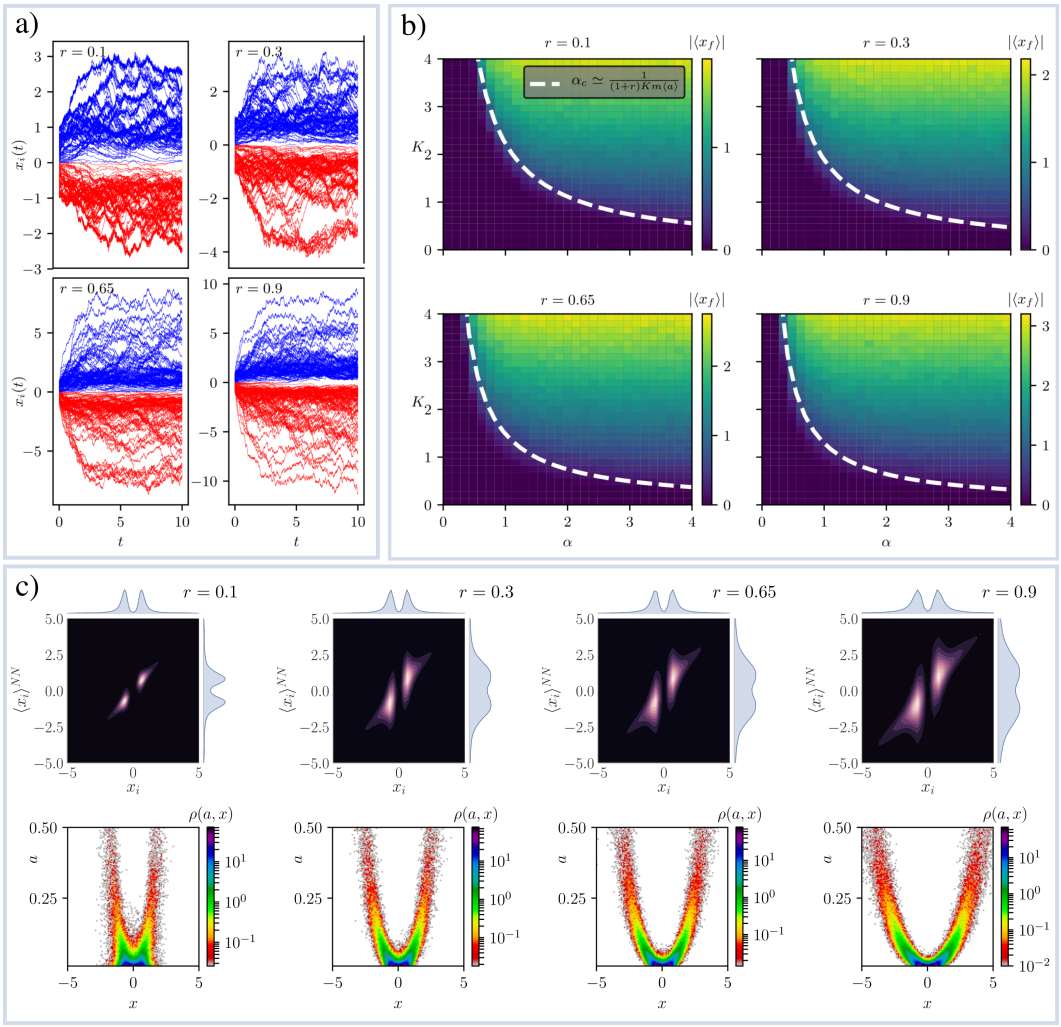

Figure B.1: Panel (a) and (b): Opinion polarization for $T = 1$ and transition to radicalization depicted in K-α space for different values of the reciprocity parameter r. The remaining parameters are identical to those in Fig. 7.2(c) and Fig. 7.3. Panel (c): Echo chambers and activity-opinion relations for different reciprocity values. The remaining parameters are set as in Fig. 7.5(b) and Fig. 7.6(a).

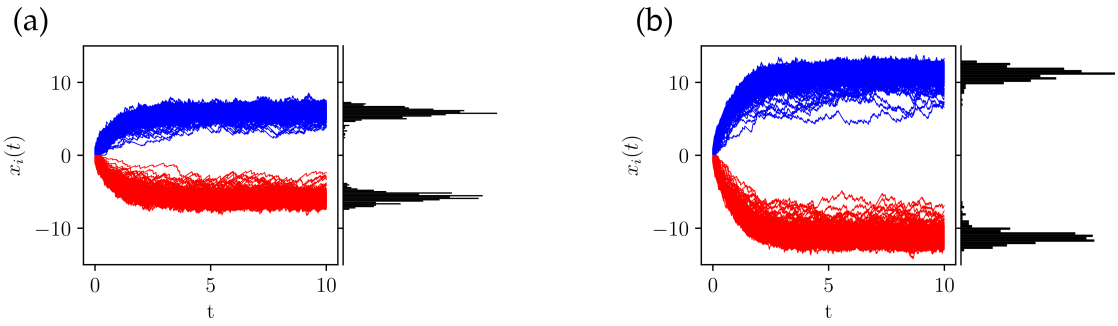

Figure B.2: Polarized opinion states and the corresponding stationary opinion distribution for a single simulation with $N = 1000$. Each agent in the system has the same activity a_c, i.e., we considered delta distributed activities with $F(a) = \delta(a - a_c)$. In panels (a) and (b) we chose $a_c = 0.1$ and $a_c = 0.2$, respectively. All other parameters were set as in Fig. 7.2(c). Note the widening of the gap between the opinion peaks for a larger value of a_c.

Twitter data

The data contrasted to our one-dimensional model introduced in Ch. 7 has been collected and analyzed in previous studies [71, 281]. The data set contains tweets on three different controversial topics of discussion: *abortion*, *Obamacare* and *guncontrol*. To generate independent data sets for each topic, we exclude users who are present in more than one data set, i.e., tweeted about more than one of the topics. Furthermore, to exclude bots, simple checks where performed in Ref. [71]. First, it was ensured that the Twitter account is older than one year (at the time of the data collection). Second, minimum and maximum thresholds for the number of tweets per day, followers and friends were implemented.

More specifically, the data sets for each topic are built by collecting tweets posted during certain events sparking increased interest in the corresponding topic (abortion, Obamacare, guncontrol). For example, the Democrat filibuster for guncontrol reforms in June 2016. The time span of data collection was one week: three days before and three days after the event. Users with less than five tweets during these periods of one week were disregarded. Finally, the numbers of users (N_u) and measured reciprocities (r) in each data set are: $N_u = 4130$ and $r = 0.69$ (*abortion*), $N_u = 4828$ and $r = 0.62$ (*Obamacare*), $N_u = 1838$ and $r = 0.61$ (*guncontrol*).

For each data set the directed follower network among users has been reconstructed [71]. In this network a directed link from node u to node v indicates that user u follows user v. The political leaning of each user is inferred from the sharing content s/he produces (via tweets) based on a ground truth of political leaning scores of news organizations (e.g. nytimes.com) established by Basky et al. [72]. In particular, each news organization is classified by a leaning score, which takes a value between 0 and 1, where a value close to 1 (0) means that the news outlet has a strongly conservative (liberal) leaning. From this classification of news outlets, the political leaning scores (opinions) of individual users are computed as follows. First, all tweets posted by user i that contain a link to an online news organization with a known political

leaning are considered. Second, the average value of these leaning scores expressed in the tweets of user i gives their opinion on a specific topic. For ease of comparison with our model results, we transformed the 0-1 scale of political leanings to one ranging from -1 to 1.

Additional multi-dimensional opinion states

Generally topics can be characterized by different values of controversialness, which yields the following generalization of Eqs. (8.4) in two dimensions:

$$\begin{aligned}\dot{x}_i^{(u)} &= -x_i^{(u)} + K\sum_j A_{ij}(t)\tanh\left[\alpha_u(x_j^{(u)} + \cos(\delta)x_j^{(v)})\right]\\ \dot{x}_i^{(v)} &= -x_i^{(v)} + K\sum_j A_{ij}(t)\tanh\left[\alpha_v(\cos(\delta)x_j^{(u)} + x_j^{(v)})\right],\end{aligned} \tag{B.1}$$

where α_u and α_v corresponds to the controversialness values of topic u and v, respectively. Below in Fig. B.3 we depict a case with $\alpha_u \neq \alpha_v$, where the opinion states with respect to each of the considered topics strongly depends on the respective controversialness, see panel (a) and (c). Due to the small value of $\alpha_u(= 0.05)$, there is a consensus with respect to topic u, while agents' opinions with respect to topic v are strongly polarized, since $\alpha_v(= 3)$ is large. This behavior is captured by the mean-field approximation, where two stable fixed points at $\mathbf{x}^* \simeq (0, \pm 1)$ arise as shown in panel (c).

As discussed in Sec. 8.3 the phase space diagram depicted in Fig. 8.3 is symmetric with respect to the line of vanishing overlaps. Additionally, we showed that the dynamics towards ideological states is reversed for negative overlaps, i.e., $\cos(\delta) < 0$ yields ideological states with negative opinion correlations. In panels (b) and (d) of Fig. B.3 such a situation is depicted for $\delta = 3\pi/4$. The emerging opinion state, as depicted for the full model [panel (b)] with $\rho(x^{(u)}, x^{(v)}) \simeq -1$, and the mean-field approximation [panel (d)] is the complementary mirrored state to the case of $\delta = \pi/4$, as shown in Fig. 8.2c.

Opinion dynamics for different levels of homophily ($T = 2$)

While in Ch. 8 we focus on cases of large homophily, both in stochastic simulations and in the mean-field analysis, here we complement our findings by simulations of Eq. (8.4) with low and intermediate values of β. The results are shown in Fig. B.4.

As in the one-dimensional model, vanishing homophily ($\beta = 0$) directly leads to a one-sided radicalized state [panel (a)], cf. Ch. 7. Increasing values of β change this picture. Although, for $\beta = 1$, the system also enters a one-sided state the process is slowed down by reinforcing interactions among like-minded peers [panel (b)]. For $\beta = 1.25$ [panel (c)] opinion polarization indeed emerges, however, in a weakened manner compared to cases of high homophily. In contrast to panel (d) showing the case of $\beta = 3$, the marginal opinion distributions depicted in panel (c) do not show a strong bimodality, which also results in a less pronounced community

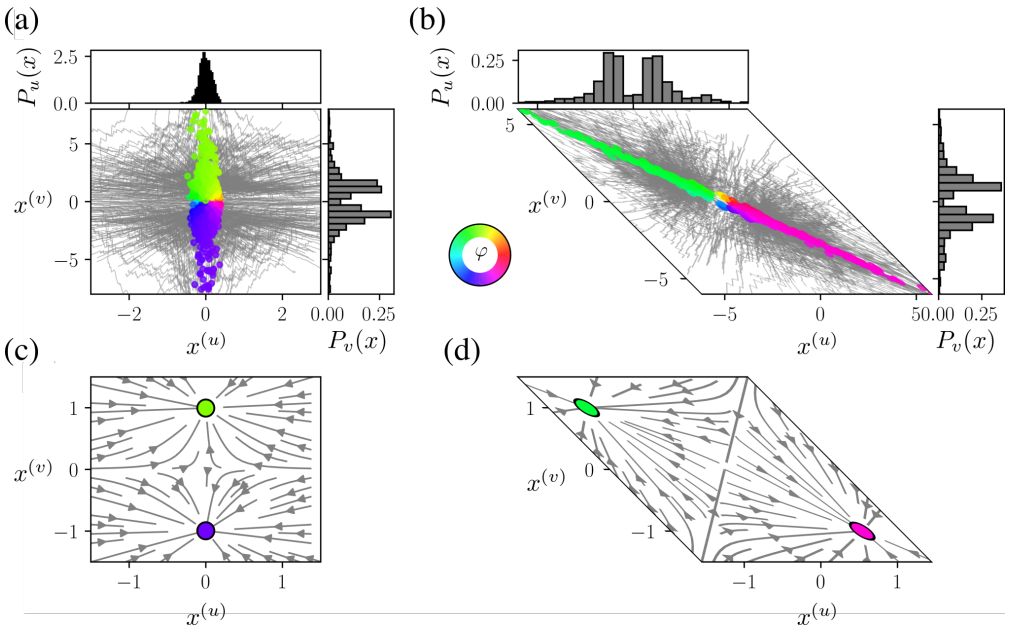

Figure B.3: Additional results of simulations of the full model (top row) and the corresponding mean-field approximations (bottom row) in two dimensions ($T = 2$). In panels (a), (c) we depict a special case in which the controversialness values (α) are different for both topics under consideration, see Eq. (B.1). In panels (b), (d) we show the resulting dynamics for a negative topic overlap $\cos(\delta) = -1/\sqrt{2}$.

structure on a network level. Additionally, for $\beta = 1.25$ the opinion trajectories (grey lines) fluctuate stronger compared to the case of $\beta = 3$, which is due to the fact that social interactions are more heterogeneous for low levels of homophily.

Numerical procedure for the generation of the phase space diagram

The stability regions shown in phase space diagram (Fig. 8.3) are computed based on the Jacobian defined in Eq. 8.6 for $T = 2$, which yieds

$$\mathbb{J}(0) = \begin{pmatrix} -1 + \Lambda\alpha & \Lambda\alpha\cos(\delta) \\ \Lambda\alpha\cos(\delta) & -1 + \Lambda\alpha \end{pmatrix}, \tag{B.2}$$

with $\Lambda = 2Km\langle a\rangle$. While the critical controversialness α_c for $T = 2$ is given by Eq. (8.11) which defines the consensus phase analytically, the regions of uncorrelated polarization and polarized ideological states need to be determined numerically. We define the phase of uncorrelated polarization by states having four stable fixed points $\mathbf{x}^*$ with $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = \{(-,+),(+,-),(+,+),(-,-)\}$. Ideological states in turn are defined by states with only two stable fixed points: (i) $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = \{(+,+),(-,-)\}$ for $\cos(\delta) > 0$, and (ii) $[\mathrm{sgn}(x^{(u)*}), \mathrm{sgn}(x^{(v)*})] = \{(+,-),(-,+)\}$ for $\cos(\delta) < 0$.

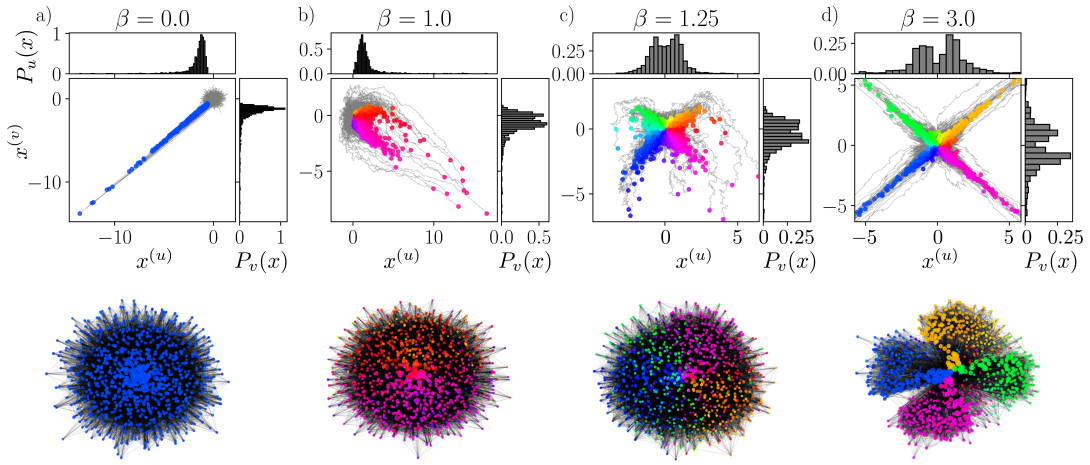

Figure B.4: Opinion dynamics and aggregated influence networks for increasing values of $\beta = [0, 1, 1.25, 3]$. The results were obtained by simulating Eqs. (8.4) for systems of $N = 1000$ agents and the following parameters: $m = 10$, $K = 3$, $\alpha = 3$, $\varepsilon = 0.01$ and $\gamma = 2.1$. The temporal networks were aggregated over 70 time steps.

The stability of each fixed point is determined numerically in two steps. First, the $\cos(\delta)$-α plane is discretized and the set of fixed points at a specific parameter combination of $\{\cos(\delta), \alpha\}$ is computed using a Newton-Raphson method [323]. In a second step, the stability of each fixed point $\mathbf{x}^*$ is determined by computing the largest eigenvalue of $\mathbb{J}(\mathbf{x}^*)$, defined in Eq. (B.2): if it is negative the corresponding fixed point is stable, otherwise it is unstable.

Additional information on ANES data

Selection procedure of analyzed questions

The 2016 American National Election Survey (ANES) [90] contains a total set of 1842 questions. Crucially, each of the 4270 respondents is assigned an individual ID. This allows us to correlate responses given by a respondent to different questions. To quantify the degree of polarization and issue alignment we compute the variances of responses to single questions and the Pearson correlation coefficients ρ between the responses to pairs of questions, respectively. In the caption of Fig. 8.6, these values are reported for the three examples of question pairs discussed in the main text.

This type of analysis requires a numerical scale for the responses. Accordingly, we first exclude all questions with free-text answers, such as “What kind of work did you do on your last regular job?". The remaining questions are of multiple-choice type, however, not all are well suited for our purpose. We only select those questions allowing us to extract the extent of approval (or disapproval) of the respondent with respect to issue, or topic, subject to the question. In particular, we use questions whose response scale allows us to quantify both the qualitative

stance (favor or oppose) and the conviction (e.g., favor a great deal, ..., neutral, ..., strongly oppose) of the respondent towards the issue, with at least a 4-point scale. Questions without such a response scale or questions which do not ask about a specific opinion, for instance, “Which of the following radio programs do you listen to regularly?" are excluded. In a final step, we exclude questions regarding political parties or presidential candidates. These selection criteria condense the 2016 ANES data set to a total of 67 questions, depicted in Fig. 8.6. The complete list of selected questions is reported below, together with the question IDs in order to locate the questions in the original data set provided by Ref. [90].

Further details on analysed questions and additional examples of correlated question pairs

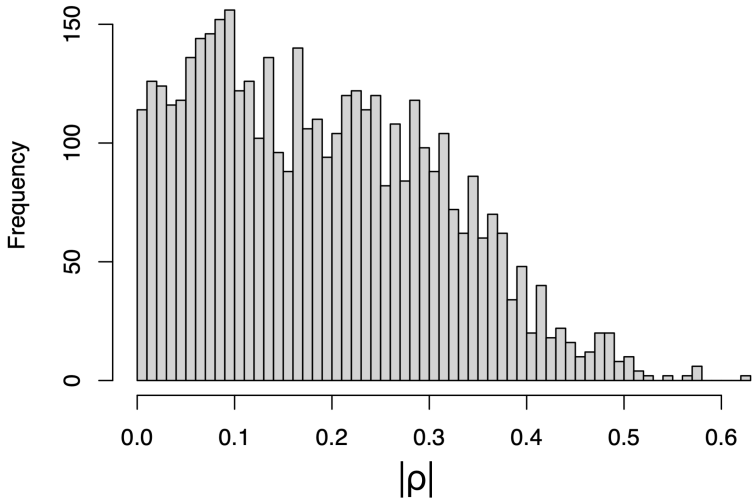

Figure B.5: Histogram of the Pearson correlation values between all the 67 analyzed questions.

Question tuple (u, v)	σ_u^2	σ_v^2	$\lvert\rho(u,v)\rvert$	p-value of ρ
"Obamacare", "voting: duty or choice"	0.586	0.646	0.032	0.032
"use of military", "free trade"	0.085	0.257	0.020	0.223
"wall with Mexico", "transgender bathroom"	0.623	0.449	0.449	0.0
"attitude towards muslims", "services to same-sex couples"	0.210	0.478	0.202	0.0
"services to same-sex couples", "transgender bathroom"	0.478	0.449	0.504	0.0
"environment regulations", "insurance plan"	0.302	0.41	0.503	0.0
"climate change action", "transgender bathroom"	0.387	0.449	0.392	0.0
"asylum for Syrian refugees", "transgender bathroom"	0.46	0.449	0.484	0.0
"asylum for Syrian refugees", "blacks should help themselves"	0.46	0.355	0.484	0.0

Table B.1: Variances of responses to single questions (σ_u^2, σ_v^2), and the corresponding Pearson correlation values between pairs of responses, $|\rho(u,v)|$, for the question combinations depicted in Fig. 8.6 and Fig. B.6.

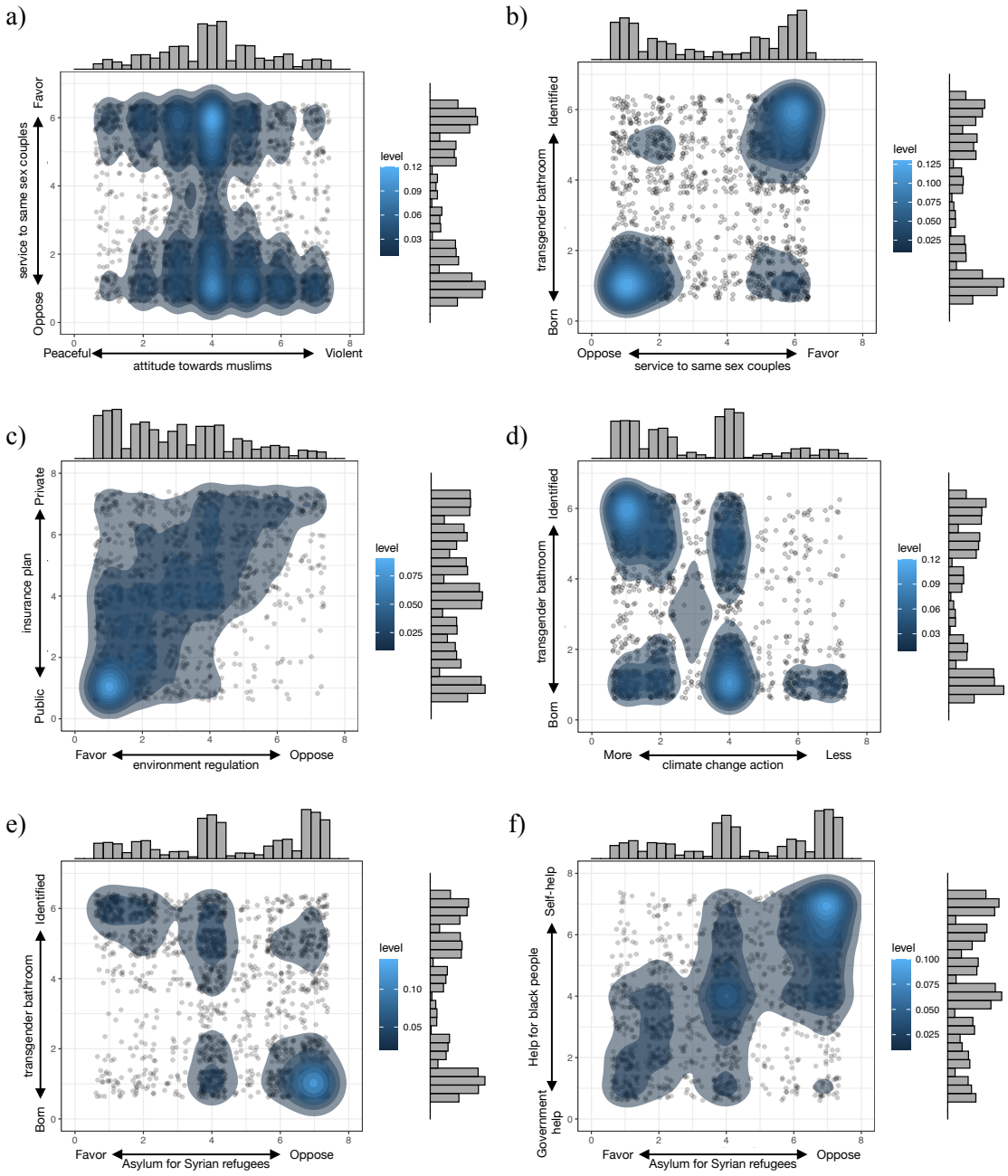

Figure B.6: Scatter plots of additional pairs of questions, complementary to those shown in Fig. 8.6.

Table B.2: Overview of all 67 analyzed questions including their abbreviated labels (left column) and the corresponding ANES IDs (right column). Blue labels correspond to the questions shown in Fig. 8.6(b)-(d) and Fig. B.6(a)-(f).

Question label	Question/Issue	ANES ID
Obamacare	Summary: favor/oppose 2010 health care law **V161114a**: IF R FAVORS THE 2010 HEALTH CARE LAW: Do you favor that [a great deal, moderately, or a little / a little, moderately, or a great deal]? **V161114b**: IF R OPPOSES THE 2010 HEALTH CARE LAW: Do you oppose that [a great deal, moderately, or a little / a little, moderately, or a great deal]?	**V161114x**
voting: duty or choice	Summary: Voting as duty or choice **V161151a**: IF R CONSIDERS VOTING A DUTY: How strongly do you feel that voting is a duty? [Very strongly, moderately strongly, or a little strongly / A little strongly, moderately strongly, or very strongly]? **V161151b**: IF R CONSIDERS VOTING A CHOICE: How strongly do you feel that voting is a choice? [Very strongly, moderately strongly, or a little strongly / A little strongly, moderately strongly, Or very strongly]?	**V161151x**
use of military	How willing should the United States be to use military force to solve international problems? [Extremely willing, very willing, moderately willing, a little willing, or not at all willing / Not at all willing, a little willing, moderately willing, very willing, or extremely willing]?	**V161154**
insurance plan	Where would you place yourself on this scale, or haven't you thought much about this? 1 (Govt insurance plan) – 7 (Private insurance plan)	**V161184**
wall with Mexico	Summary: Build wall with Mexico **V161196**: Do you favor, oppose, or neither favor nor oppose building a wall on the U.S. border with Mexico? **V161196a**: IF R FAVORS BUILDING A WALL ON THE U.S. BORDER WITH MEXICO / IF R OPPOSES BUILDING A WALL ON THE U.S. BORDER WITH MEXICO: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that[a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V161196x**
help for black people	Where would you place yourself on this scale, or haven't you thought much about this? 1 (Govt should help Blacks) – 7 (Blacks should help themselves)	**V161198**
environment regulations	Where would you place yourself on this scale, or haven't you thought much about this? 1 (Regulate business to protect the environment and create jobs) – 7 (No regulation because it will not work and will cost jobs)	**V161201**

asylum for Syrian refugees	Summary: Allow Syrian refugees **V161214**: Do you favor, oppose, or neither favor nor oppose allowing Syrian refugees to come to the United States? **V161214a**: IF R FAVORS ALLOWING SYRIAN REFUGEES TO COME TO THE U.S. / IF R OPPOSES ALLOWING SYRIAN REFUGEES TO COME TO THE U.S.: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V161214x**
climate change action	Summary: Govt action about rising temperatures **V161224**: Do you think the federal government should be doing more about rising temperatures, should be doing less, or is it currently doing the right amount? **V161224a**: IF R SAYS GOVERNMENT SHOULD DO MORE ABOUT RISING TEMPERATURES / IF R SAYS GOVERNMENT SHOULD DO LESS ABOUT RISING TEMPERATURES: Should it be doing a great deal [more/less], a moderate amount [more/less], or a little [more/less]? / Should it be doing a little [more/less], a moderate amount [more/less], or a great deal [more/less]?	**V161225x**
free trade	Summary: Favor/oppose free trade agreements **V162176**: Do you favor, oppose, or neither favor nor oppose the U.S. making free trade agreements with other countries? **V162176a**: How strongly do you [favor/oppose] it?	**V162176x**
transgender bathroom	Summary: Transgender policy **V161228**: Should transgender people – that is, people who identify themselves as the sex or gender different from the one they were born as – have to use the bathrooms of the gender they were born as, or should they be allowed to use the bathrooms of their identified gender? **V161228a**: IF R OPINION ON TRANSGENDER USE OF RESTROOMS OF IDENTIFIED GENDER IS NOT DK/RF: How strongly do you feel about that? [Very strongly, moderately strongly, or slightly strongly / Slightly strongly, moderately strongly or very strongly]?	**V161228x**
attitude towards muslims	Where would you rate Muslims in general on this scale? 1 (Peaceful) - 7 (Violent)	**V162353**

service to same sex couples	Summary: Services to same sex couples **V161227**: Do you think business owners who provide wedding-related services should be allowed to refuse services to same-sex couples if same-sex marriage violates their religious beliefs, or do you think business owners should be required to provide services regardless of a couple's sexual orientation? **V161227a**: IF R OPINION ON REFUSING WEDDING SERVICES TO SAME-SEX COUPLES IS NOT DK/RF: How strongly do you feel that way? [Very strongly, moderately, or a little / A little, moderately, or very strongly]?	**V161227x**
government services	Where would you place yourself on this scale, or haven't you thought much about this? 1 (Govt should provide many fewer services) – 7 (Govt should provide many more services)	**V161178**
defense spending	Where would you place yourself on this scale, or haven't you thought much about this? 1 (Govt should decrease defense spending) – 7 (Govt should increase spending)	**V161181**
gun access	How important is this issue [gun access] to you personally? [Extremely important, very important, somewhat important, not too important, or not important at all / Not important at all, not too important, somewhat important, very important, or extremely important]?	**V161188**
birthright	Summary: Birthright citizenship **V161193**: Some people have proposed that the U.S. Constitution should be changed so that the children of unauthorized immigrants do not automatically get citizenship if they are born in this country. Do you favor, oppose, or neither favor nor oppose this proposal? **V161193a**: IF R FAVORS CHANGING CONSTITUTION - US-BORN CHILDREN OF ILLEGAL IMMIGRANTS / IF R FAVORS CHANGING CONSTITUTION - US-BORN CHILDREN OF ILLEGAL IMMIGRANTS: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V161194x**
immigrant children	Summary: Children brought illegally **V161195**: What should happen to immigrants who were brought to the U.S. illegally as children and have lived here for at least 10 years and graduated high school here? Should they be sent back where they came from, or should they be allowed to live and work in the United States? **V161195a**: IF R OPINION ON ILLEGAL IMMIGRANT CHILDREN RAISED IN U.S. IS NOT DK/RF: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, a great deal]?	**V161195x**

importance of english	How important do you think it is that everyone in the United States learn to speak English? Very important, somewhat important, not very important, or not at all important ?	**V161197**
affirmative action	Summary: Favor or oppose affirmative action in universities **V161204**: Do you favor, oppose, or neither favor nor oppose allowing universities to increase the number of black students studying at their schools by considering race along with other factors when choosing students? **V161204a**: IF R FAVORS AFFIRMATIVE ACTION AT UNIVERSITIES: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? **V161204b**: IF R OPPOSES AFFIRMATIVE ACTION AT UNIVERSITIES: Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V161204x**
fight ISIS	Summary: Send troops to fight ISIS **V161213**: Do you favor, oppose, or neither favor nor oppose the U.S. sending ground troops to fight Islamic militants, such as ISIS, in Iraq and Syria? **V161213a**: IF R FAVORS SENDING U.S. GROUND TROUPS TO FIGHT ISLAMIC MILITANTS LIKE ISIS / IF R OPPOSES SENDING U.S. GROUND TROUPS TO FIGHT ISLAMIC MILITANTS LIKE ISIS: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great Deal]?	**V161213x**
parental leave	Summary: Require employers to offer paid leave to new parents **V161226**: Do you favor/oppose, or neither favor nor oppose requiring employers to offer paid leave to parents of new children? **V161226a**: IF R FAVORS REQUIRING EMPLOYERS TO OFFER PAID LEAVE FOR NEW CHILDREN / IF R OPPOSES REQUIRING EMPLOYERS TO OFFER PAID LEAVE FOR NEW CHILDREN : Do you favor that a great deal, a moderate amount, or a little ? / Do you oppose that a great deal, a moderate amount, or a little?	**V161226x**
protection of gay/lesb	Summary: Laws to protect gays and lesbians against job discrim **V161229**: Do you favor or oppose laws to protect gays and lesbians against job discrimination? **V161229a**: IF R FAVORS PROTECTING GAYS AND LESBIANS AGAINST JOB DISCRIMINATION/ IF R OPPOSES PROTECTING GAYS AND LESBIANS AGAINST JOB DISCRIMINATION: [Do you favor such laws strongly or not strongly? / Do you oppose such laws strongly or not Strongly?]	**V161229x**

abortion	There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view? 1 (By law, abortion should never be permitted.), 2 (By law, only in case of rape, incest, or woman's life in danger.), 3 (By law, for reasons other than rape, incest, or woman's life in danger if needed established) 4 (By law, abortion as a matter of personal choice)	**V161232**
death penalty	Summary: Favor or oppose death penalty **V161233**: Do you favor or oppose the death penalty for persons convicted of murder? **V161233a**: IF R FAVORS DEATH PENALTY FOR PERSONS CONVICTED OF MURDER / IF R OPPOSES DEATH PENALTY FOR PERSONS CONVICTED OF MURDER: Do you [favor / oppose] the death penalty for persons convicted of murder strongly or not strongly?	**V161233x**
religious services	IF R ATTENDS RELIGIOUS SERVICES: Do you go to religious services [every week, almost every week, once or twice a month, a few times a year, or never/ never, a few times a year, once or twice a month, almost every week, or every week]?	**V161245**
rough-up protesters	When protestors get 'roughed up' for disrupting political events, how much do they generally deserve what happens to them?	**V161343**
feminism	How well does the term feminist' describe you?	**V161346**
language sensitivity	Some people think that the way people talk needs to change with the times to be more sensitive to people from different backgrounds. Others think that this has already gone too far and many people are just too easily offended. Which is closer to your opinion?	**V161362**
woman not appreciate men	'Most women fail to appreciate fully all that men do for them.' (Do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with this statement?)	**V161508**
countries like America	'The world would be a better place if people from other countries were more like Americans.' Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?	**V162123**
american flag	Summary: How good/bad does R feel to see American flag **V162125**: IF R SEEING THE AMERICAN FLAG MAKES R FEEL GOOD / IF R SEEING THE AMERICAN FLAG MAKES R FEEL BAD: Does it make you feel [extremely good, moderately good, or a little good / a little good, moderately good, or extremely good]? / Does it make you feel [extremely bad, moderately bad, or a little bad / a little bad, moderately bad, or extremely bad]?	**V162125x**

vaccines	Summary: Favor/oppose vaccines in schools **V162147**: IF R FAVORS REQUIRING VACCINATION IN ORDER FOR CHILDREN TO ATTEND SCHOOL / IF R OPPOSES REQUIRING VACCINATION IN ORDER FOR CHILDREN TO ATTEND SCHOOL: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V162147x**
equal pay	Summary: Favor/oppose equal pay for men and women **V162150**: IF FAVORS REQUIRING EMPLOYERS TO PAY MEN AND WOMEN SAME FOR THE SAME WORK/ IF OPPOSES REQUIRING EMPLOYERS TO PAY MEN AND WOMEN SAME FOR THE SAME WORK: Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Do you oppose that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V162150x**
support for israel	Summary: How much should U.S. support Israelis **V162155a**: In the conflict between Palestinians and Israelis, how much should the United States support the Palestinians? [A great deal, a lot, a moderate amount, a little, or not at all / Not at all, a little, a moderate amount, a lot, or a great deal]? **V162155b**: In this conflict, how much should the United States support the Israelis? [A great deal, a lot, a moderate amount, a little, or not at all / Not at all, a little, a moderate amount, a lot, or a great deal]?	**V162155x**
immigration	Do you think the number of immigrants from foreign countries who are permitted to come to the United States to live should be increased/decreased?	**V162157**
threat of China's military	Do you think China's military is [a major threat to the security of the United States, a minor threat, or not a threat / not a threat, a minor threat, or a major threat to the security of the United States] ?	**V162159**
worries about terrorists	How worried are you that the United States will experience a terrorist attack in the near future?	**V162160**
free thinkers	'Our country needs free thinkers who will have the courage to defy traditional ways, even if this upsets many people.' Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?	**V162168**
honor forefathers	'Our country would be great if we honor the ways of our forefathers, do what the authorities tell us to do, and get rid of the 'rotten apples' who are ruining everything.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?)	**V162169**

strong leader	'What our country really needs is a strong, determined leader who will crush evil and take us back to our true path.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?)	**V162170**
business regulation	How much government regulation of business is good for society? [A great deal, a lot, a moderate amount, a little, or none at all / None at all, a little, a moderate amount, a lot, or a great deal]?	**V162186**
spending on healthcare	Summary: Increase/decrease gov spending for health care **V162193**: Do you favor an increase, decrease, or no change in government spending to help people pay for health insurance when they can't pay for it all themselves? **V162193a**: Should it increase [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]? / Should it decrease [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?	**V162193x**
traditional family	'This country would have many fewer problems if there were more emphasis on traditional family ties.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?)	**V162210**
help blacks	'Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.' Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?	**V162211**
have a say in govt	'People like me don't have any say about what the government does.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?)	**V162216**
understand politics	How often do politics and government seem so complicated that you can't really understand what's going on? [Always, most of the time, about half the time, some of the time, or never / Never, some of the time, about half the time, most of the time, or always?	**V162217**
influence of money on elections	(In your view, how often do the following things occur in this country's elections?) Rich people buy elections [All of the time , most of the time, about half of the time, some of the time, never / Never/ some of the time, about half of the time, most of the time, or all of the time?]	**V162220**

hispanics in politics	How important is it that more Hispanics be elected to political office? [Extremely important, very important, moderately important, a little important, or not important at all / Not important at all, a little important, moderately important, very important, or extremely important]?	**V162221**
woman should work	Summary: Better if man works and woman takes care of home **V162230**: Do you think it is better, worse, or makes no difference for the family as a whole if the man works outside the home and the woman takes care of the home and family? **V162230a**: IF R SAYS IT IS BETTER FOR THE MAN TO WORK AND THE WOMAN TO STAY AT HOME: Is it [much better, somewhat better, or slightly better / slightly better, somewhat better or much better]? **V162230b**: IF R SAYS IT IS WORSE FOR THE MAN TO WORK AND THE WOMAN TO STAY AT HOME: Is it [much worse, somewhat worse, or slightly worse / slightly worse, somewhat worse or much worse]?	**V162230x**
attention to women discr	Summary: How much attn media should pay to discrim against Women **V162231**: Should the news media pay more attention to discrimination against women, less attention, or the same amount of attention they have been paying lately? **V162231a**: IF THE NEWS MEDIA SHOULD PAY MORE ATTENTION TO DISCRIMINATION AGAINST WOMEN: Should the media pay [a great deal more attention, somewhat more attention, or a little more attention / a little more attention, somewhat more attention, or a great deal more attention]? **V162231b**: IF THE NEWS MEDIA SHOULD PAY LESS ATTENTION TO DISCRIMINATION AGAINST WOMEN: Should the media pay [a great deal less attention, somewhat less attention, or a little less attention / a little less attention, somewhat less attention, or a great deal less attention]?	**V162231x**
women's rights	When women demand equality these days, how often are they actually seeking special favors? [Always, most of the time, about half the time, some of the time, or never / Never, some of the time, about half the time, most of the time, or always ?	**V162232**
preferentially hire blacks	Summary: Favor preferential hiring and promotion of blacks **V162238**: What about your opinion – are you for or against preferential hiring and promotion of blacks? **V162238a**: IF R IS FOR PREFERENTIAL HIRING AND PROMOTION FOR BLACKS: Do you favor preference in hiring and promotion strongly or not strongly? **V162238b**: IF R IS AGAINST PREFERENTIAL HIRING AND PROMOTION FOR BLACKS: Do you oppose preference in hiring and promotion strongly or not strongly?	**V162238x**

importance of equality	'This country would be better off if we worried less about how equal people are.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat, or agree strongly] with this statement?)	**V162244**
obama is / is not muslim	Summary: Barack Obama is/isn't Muslim **V162255**: Is Barack Obama a Muslim, or is he not a Muslim? **V162255a**: IF R SAYS THAT BARACK OBAMA IS A MUSLIM OR SAYS THAT BARACK OBAMA IS NOT A MUSLIM: How sure are you about that? [Extremely sure, very sure, moderately sure, a little sure, or not at all sure / Not at all sure, a little sure, moderately sure, very sure, or extremely sure]?	**V162255x**
strong leader bending rules	'Having a strong leader in government is good for the United States even if the leader bends the rules to get things done.' (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat or agree strongly]?)	**V162263**
minorities should adopt traditions	Now thinking about minorities in the United States. Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat or agree strongly] with the following statement? 'Minorities should adapt to the customs and traditions of the United States'	**V162266**
immigrants good for economy	And now thinking specifically about immigrants. (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly /disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat or agree strongly] with the following statement?) 'Immigrants are generally good for America's economy.'	**V162268**
immigrants increase crime	(Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly /disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat or agree strongly] with the following statement?) 'Immigrants increase crime rates in the United States.'	**V162270**
born in US	Some people say that the following things are important for being truly American. Others says they are not important. How important do you think the following is for being truly American... [very important, fairly important, not very important, or not important at all / not important at all, not very important, fairly important or very important]? To have been born in the United States	**V162271**
corruption in politics	How widespread do you think corruption such as bribe taking is among politicians in the United States: [Very widespread, quite widespread, not very widespread, or it hardly happens at all / It hardly happens at all, is not very widespread, quite widespread, or very widespread]?	**V162275**

reduce income differences	Please say to what extend you agree or disagree with the following statement: 'The government should take measures to reduce differences in income levels'. (Do you [agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly / disagree strongly, disagree somewhat, neither agree nor disagree, agree somewhat or agree strongly]?)	**V162276**
torture of terrorists	Summary: Favor/oppose torture for suspected terrorists **V162295**: Do you favor, oppose, or neither favor nor oppose the U.S. government torturing people who are suspected of being terrorists, to try to get information? **V162295a**: IF R FAVORS USE OF TORTURE AGAINST SUSPECTED TERRORISTS: Do you favor that [a great deal, moderately, or a little / a little, moderately, or a great deal]? **V162295b**: IF R OPPOSES USE OF TORTURE AGAINST SUSPECTED TERRORISTS: Do you oppose that [a great deal, moderately, or a little / a little, moderately, or a great deal]?	**V162295x**
whites lazy	Where would you rate Whites in general on this scale? 1 (Hard-working) – 7 (Lazy)	**V162345**
christians patriotic	Where would you rate Christians in general on this scale? 1 (Patriotic) – 7 (Unpatriotic)	**V162356**
discrimination against blacks	How much discrimination is there in the United States today against each of the following groups? Blacks	**V162357**
discrimination against women	How much discrimination is there in the United States today against each of the following groups? Women	**V162362**

Bibliography

[1] Andrés Chacoma and Damián H. Zanette. Opinion formation by social influence: From experiments to modeling. *PLOS ONE*, 10(10):1–16, 2015.

[2] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis P. Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 2013.

[3] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4):2, 2017.

[4] Winter A. Mason, Frederica R. Conrey, and Eliot R. Smith. Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. *Personality and Social Psychology Review*, 11(3):279–300, 2007.

[5] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003.

[6] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[7] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, page 33–42, New York, NY, USA, 2012.

[8] Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. Accelerating dynamics of collective attention. *Nature communications*, 10(1):1–9, 2019.

[9] Seth A. Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy J. Lin. Information network or social network?: the structure of the twitter follow graph. In *WWW (Companion Volume)*, pages 493–498, 2014.

[10] Marijke De Veirman, Veroline Cauberghe, and Liselot Hudders. Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude. *International journal of advertising*, 36(5):798–828, 2017.

[11] Shanto Iyengar. *Is anyone responsible?: How television frames political issues*. University of Chicago Press, 1994.

[12] Marcel Broersma and Todd Graham. Twitter as a news source: How dutch and british newspapers used tweets in their news coverage, 2007–2011. *Journalism practice*, 7(4): 446–464, 2013.

[13] Joshua A. Tucker, Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. From liberation to turmoil: Social media and democracy. *Journal of democracy*, 28(4):46–59, 2017.

[14] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th International Conference on World Wide Web*, pages 355–356, 2015.

[15] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi. Polarization of the vaccination debate on facebook. *Vaccine*, 36(25):3606 – 3612, 2018.

[16] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.

[17] Georg Simmel. Der Streit. *Soziologie. Untersuchungen über die Formen der Vergesellschaftung*, 11:284–382, 1908.

[18] Paul DiMaggio, John Evans, and Bethany Bryson. Have american's social attitudes become more polarized? *American Journal of Sociology*, 102(3):690–755, 1996.

[19] David Lewis. *Convention: A philosophical study*. John Wiley & Sons, 2008.

[20] Andrea Baronchelli. The emergence of consensus: a primer. *Royal Society open science*, 5(2):172189, 2018.

[21] John Urry. Climate change and society. In *Why the social sciences matter*, pages 45–59. Springer, 2015.

[22] Robert J. Smith, Robert D.J. Muir, Matthew J. Walpole, Andrew Balmford, and Nigel Leader-Williams. Governance and the loss of biodiversity. *Nature*, 426(6962):67–70, 2003.

[23] A. Spinelli and G. Pellino. Covid-19 pandemic: perspectives on an unfolding crisis. *BJS (British Journal of Surgery)*, 107(7):785–787, 2020.

[24] Paul G. Harris. Collective action on climate change: The logic of regime failure. *Natural Resources Journal*, 47(1):195–224, 2007.

[25] David Garcia, Adiya Abisheva, Simon Schweighofer, Uwe Serdült, and Frank Schweitzer. Ideological and temporal components of network polarization in online political participatory media. *Policy & Internet*, 7(1):46–79, 2015.

[26] Cass R. Sunstein. *Why societies need dissent*, volume 9. Harvard University Press, 2005.

[27] Nolan McCarty, Keith T. Poole, and Howard Rosenthal. *Polarized America: The dance of ideology and unequal riches*. MIT Press, 2016.

[28] Stathis N. Kalyvas. *The logic of violence in civil war*. Cambridge University Press, 2006.

[29] Robyn Torok. Developing an explanatory model for the process of online radicalisation and terrorism. *Security Informatics*, 2(1):1–10, 2013.

[30] Eugene Burnstein and Amiram Vinokur. What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology*, 11(5):412 – 426, 1975.

[31] Eugene Burnstein and Amiram Vinokur. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, 13(4):315 – 332, 1977.

[32] Ronald L. Akers, Marvin D. Krohn, Lonn Lanza-Kaduce, and Marcia Radosevich. Social learning and deviant behavior: A specific test of a general theory. *American Sociological Review*, 44(4):636–655, 1979.

[33] Leon Festinger, Stanley Schachter, and Kurt Back. Social pressures in informal groups; a study of human factors in housing. 1950.

[34] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5): 992–1026, 1992.

[35] Fritz Heider. Attitudes and cognitive organization. *The Journal of Psychology*, 21(1): 107–112, 1946.

[36] Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1957.

[37] Joshua M. Epstein. *Generative social science: Studies in agent-based computational modeling*, volume 13. Princeton University Press, 2006.

[38] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

[39] Eric Bertin. *A concise introduction to the statistical physics of complex systems*. Springer Science & Business Media, 2011.

[40] Richard C. Tolman. *The principles of statistical mechanics*. Courier Corporation, 1979.

[41] Ralph H. Fowler. *Statistical thermodynamics*. CUP Archive, 1939.

[42] Paul H. E. Tiesinga. *From Neuron to Brain: Statistical Physics of the Nervous System*, pages 99–112. Springer US, Boston, MA, 2003. ISBN 978-1-4615-0207-4.

[43] Guy Sella and Aaron E. Hirsh. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences*, 102(27):9541–9546, 2005.

[44] Lambert A. J. Quetelet. *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*. 1869.

[45] Georg G. Iggers. Further remarks about early uses of the term "social science". *Journal of the History of Ideas*, 20(3):433–436, 1959.

[46] Paolo Moretti, Andrea Baronchelli, Michele Starnini, and Romualdo Pastor-Satorras. Generalized voter-like models on heterogeneous networks. In *Dynamics On and Of Complex Networks, Volume 2*, pages 285–300. Springer, 2013.

[47] Luc Steels. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332, 1995.

[48] LF Henderson. The statistics of crowd fluids. *Nature*, 229(5284):381–383, 1971.

[49] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[50] Serge Galam. Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach. *Journal of Mathematical Psychology*, 30(4):426 – 434, 1986.

[51] Frank Schweitzer. Sociophysics. *Physics today*, 71(2):40–46, 2018.

[52] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1): 253–258, 1925.

[53] Maciej Lewenstein, Andrzej Nowak, and Bibb Latané. Statistical mechanics of social impact. *Phys. Rev. A*, 45:763–776, 1992.

[54] Sven Banisch and Eckehard Olbrich. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, 43(2):76–103, 2019.

[55] Dirk Helbing. *Social self-organization. Agent-based simulations and experiments to study emergent social behavior*. Understanding Complex Systems. Springer, Heidelberg, 2012.

[56] Gerben A. Van Kleef. How emotions regulate social life: The emotions as social information (easi) model. *Current Directions in Psychological Science*, 18(3):184–188, 2009.

[57] Steven T Piantadosi. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21(5):1112–1130, 2014.

[58] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.

[59] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2:469, 2012.

[60] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009.

[61] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.

[62] Mark E. Newman. The structure and function of complex networks. *SIAM review*, 45(2): 167–256, 2003.

[63] John Scott. Social network analysis. *Sociology*, 22(1):109–127, 1988.

[64] Jacob Levy Moreno. Who shall survive?: A new approach to the problem of human interrelations. 1934.

[65] Peter V. Marsden. Network data and measurement. *Annual Review of Sociology*, 16(1): 435–463, 1990.

[66] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[67] Mark E. Newman, Steven H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, 2001.

[68] Albert L. Barabási, Hawoong Jeong, Zoltan Néda, Erzsébet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3):590 – 614, 2002.

[69] Göran Melin and Olle Persson. Studying research collaboration using co-authorships. *Scientometrics*, 36(3):363–377, 1996.

[70] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, 2004.

[71] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, pages 913–922, Switzerland, April 2018.

[72] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[73] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[74] Piotr Fronczak. *Scale-Free Nature of Social Networks*. Springer New York, 2018.

[75] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Scientific reports*, 3(1):1–6, 2013.

[76] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[77] Noah E. Friedkin and Eugene C. Johnsen. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Structural Analysis in the Social Sciences. Cambridge University Press, 2011.

[78] Ping-Ping Li, Da-Fang Zheng, and Pak M. Hui. Dynamics of opinion formation in a small-world network. *Phys. Rev. E*, 73:056128, 2006.

[79] Dina Mistry, Qian Zhang, Nicola Perra, and Andrea Baronchelli. Committed activists and the reshaping of status-quo social consensus. *Phys. Rev. E*, 92:042805, 2015.

[80] Jae K. Shin and Jan Lorenz. Tipping diffusivity in information accumulation systems: more links, less consensus. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(06):P06005, jun 2010.

[81] John R. P. French Jr. A formal theory of social power. *Psychological review*, 63(3):181–94, 1956.

[82] Frank Harary. Studies in social power. In D. Cartwright, editor, *A criterion for unanimity in French's theory of social power.*, pages 168–182. University Michigan, 1959.

[83] Morris H. Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[84] Robert P. Abelson. Mathematical models of the distribution of attitudes under controversy. In Norman Frederiksen and Harold Gulliksen, editors, *Contributions to Mathematical Psychology*, pages 142–60. Rinehart Winston, New York, 1964.

[85] Wesley Cota, Silvio C. Ferreira, Romualdo Pastor-Satorras, and Michele Starnini. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1):35, 2019.

[86] R. Kelly Garrett. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2):265–285, 2009.

[87] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. Echo chambers on social media: A comparative analysis, 2020. URL https://arxiv.org/abs/2004.09603. arXiv preprint.

[88] Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, Sep 2020. ISSN 2432-2725. doi: 10.1007/s42001-020-00084-7. URL http://dx.doi.org/10.1007/s42001-020-00084-7.

[89] Xin Wang, Antonio D. Sirianni, Shaoting Tang, Zhiming Zheng, and Feng Fu. Public discourse and social network echo chambers driven by socio-cognitive biases. *Phys. Rev. X*, 10:041042, 2020.

[90] The American National Election Studies (www.electionstudies.org). These materials are based on work supported by the National Science Foundation under grant numbers SES 1444721, 2014-2017, the University of Michigan, and Stanford University.

[91] Jan Lorenz. Opinion dynamics and collective decisions: Procedures, behavior, and systems dynamics, 2019. URL https://gepris.dfg.de/gepris/projekt/265108307/ergebnisse.

[92] Jan Lorenz. Modeling the evolution of ideological landscapes through opinion dynamics. In Wander Jager, Rineke Verbrugge, Andreas Flache, Gert de Roo, Lex Hoogduin, and Charlotte Hemelrijk, editors, *Advances in Social Simulation 2015*, pages 255–266, Cham, 2017. Springer International Publishing.

[93] John R. Zaller. *The mainstream and polarization effects*, page 97–117. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, 1992.

[94] Peter K. Hatemi, Sarah E. Medland, Robert Klemmensen, Sven Oskarsson, Levente Littvay, Christopher T. Dawes, Brad Verhulst, Rose McDermott, Asbjørn Sonne Nørgaard, Casey A. Klofstad, et al. Genetic influences on political ideologies: Twin analyses of 19 measures of political ideologies from five democracies and genome-wide findings from three populations. *Behavior genetics*, 44(3):282–294, 2014.

[95] Jan Lorenz. Zur methode der agenten-basierten Simulation in der Politikwissenschaft am Beispiel von Meinungsdynamik und Parteienwettstreit. In Thomas Bräuninger, André Bächtiger, and Susumu Shikano, editors, *Jahrbuch für Handlungs- und Entscheidungstheorie*, pages 31–58, Wiesbaden, 2012. VS Verlag für Sozialwissenschaften.

[96] Thomas C. Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[97] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.

[98] Igor Tchappi Haman, Vivient C. Kamla, Stéphane Galland, and Jean C. Kamgang. Towards an multilevel agent-based model for traffic simulation. *Procedia Computer Science*, 109:887 – 892, 2017.

[99] Dirk Helbing and Péter Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51:4282–4286, 1995.

[100] Bàrbara Llacay and Gilbert Peffer. Using realistic trading strategies in an agent-based stock market model. *Computational and Mathematical Organization Theory*, 24(3): 308–350, 2018.

[101] Elmar Kiesling, Markus Günther, Christian Stummer, and Lea M. Wakolbinger. Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research*, 20(2):183–230, 2012.

[102] Joshua M. Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.

[103] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99:7280–7287, 2002.

[104] Thomas C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[105] Dandan Li, Dun Han, Jing Ma, Mei Sun, Lixin Tian, Timothy Khouw, and H. Eugene Stanley. Opinion dynamics in activity-driven networks. *EPL (Europhysics Letters)*, 120 (2):28002, 2017.

[106] Jalil Hasanyan, Lorenzo Zino, Daniel Alberto Burbano Lombana, Alessandro Rizzo, and Maurizio Porfiri. Leader-follower consensus on activity-driven networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2233): 20190485, 2020.

[107] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.*, 124: 048301, 2020.

[108] Serge Galam and Frans Jacobs. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 381: 366 – 376, 2007.

[109] Michael Macy, James Kitts, Andreas Flache, and Steve Benard. Polarization in dynamic networks: A hopfield model of emergent structure. pages 162–173, 01 2003.

[110] Diemo Urbig and Robin Malitz. Dynamics of structured attitudes and opinions. In *Third Conference of the European Social Simulation Association*, pages 5–8, 2005.

[111] Santo Fortunato, Vito Latora, Alessandro Pluchino, and Andrea Rapisarda. Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, 16(10):1535–1551, 2005.

[112] Delia Baldassarri and Peter Bearman. Dynamics of political polarization. *American Sociological Review*, 72(5):784–811, 2007.

[113] Jan Lorenz. *Fostering Consensus in Multidimensional Continuous Opinion Dynamics under Bounded Confidence*, pages 321–334. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[114] Sylvie Huet and Deffuant Guillaume. Openness leads to opinion stability and narrowness to volatility. *Advances in Complex Systems*, 13(03):405–423, 2010.

[115] Andreas Flache and Michael W. MACY. Small worlds and cultural polarization. *The Journal of Mathematical Sociology*, 35(1-3):146–176, 2011.

[116] Philip E. Converse. The nature of belief systems in mass publics (1964). *Critical Review*, 18(1-3):1–74, 2006.

[117] Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2): 408–446, 2008.

[118] Daniel DellaPosta, Yongren Shi, and Michael Macy. Why do liberals drink lattes? *American Journal of Sociology*, 120(5):1473–1511, 2015.

[119] Richard A. Holley and Thomas M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663, 1975.

[120] Cass R. Sunstein. The law of group polarization. *Journal of Political Philosophy*, 10(2): 175–195, 2002.

[121] Michael Mäs and Andreas Flache. Differentiation without distancing. Explaining bipolarization of opinions without negative influence. *PLOS ONE*, 8(11):1–17, 11 2013.

[122] André C.R. Martins. Trust in the coda model: Opinion dynamics and the reliability of other agents. *Physics Letters A*, 377(37):2333 – 2339, 2013.

[123] André C. R. Martins. Continuous Opinions And Discrete Actions In Opinion Dynamics Problems. *International Journal of Modern Physics C (IJMPC)*, 19(04):617–624, 2008.

[124] André C.R. Martins. Discrete opinion models as a limit case of the coda model. *Physica A: Statistical Mechanics and its Applications*, 395:352 – 357, 2014.

[125] Raymond S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220, 1998.

[126] Jonathan Baron. Beliefs about thinking. *Informal reasoning and education*, pages 169–186, 1991.

[127] Harold H. Kelley. The warm-cold variable in first impressions of persons. *Journal of Personality*, 18(4):431–439, 1950.

[128] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[129] Paul F. Lazarsfeld, Robert K. Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[130] Matthijs Kalmijn. Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, 24(1):395–421, 1998.

[131] Wesley Shrum, Neil H. Cheek, and Saundra MacD. Hunter. Friendship in school: Gender and racial homophily. *Sociology of Education*, 61(4):227–239, 1988.

[132] James R. Lincoln and Jon Miller. Work and friendship ties in organizations: A comparative analysis of relation networks. *Administrative Science Quarterly*, 24(2):181–199, 1979.

[133] David Knoke. Networks of political action: Toward theory construction. *Social Forces*, 68(4):1041–1063, 1990.

[134] Lois M. Verbrugge. The structure of adult friendship choices. *Social Forces*, 56(2): 576–597, 1977.

[135] Robert R. Huckfeldt, John Sprague, James H. Kuklinski, and Robert S. Wyer. *Citizens, Politics and Social Communication: Information and Influence in an Election Campaign*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, 1995.

[136] Amiram Vinokur, Yaacov Trope, and Eugene Burnstein. A decision-making analysis of persuasive argumentation and the choice-shift effect. *Journal of Experimental Social Psychology*, 11(2):127 – 148, 1975.

[137] Verlin B. Hinsz and James H. Davis. Persuasive arguments theory, group polarization, and choice shifts. *Personality and Social Psychology Bulletin*, 10(2):260–268, 1984.

[138] David G. Myers. Discussion-induced attitude polarization. *Human Relations*, 28(8): 699–714, 1975.

[139] Anatol Rapoport and William J. Horvath. A study of a large sociogram. *Behavioral science*, 6(4):279–291, 1961.

[140] Alden S. Klovdahl, John J. Potterat, Donald E. Woodhouse, John B. Muth, Stephen Q. Muth, and William W. Darrow. Social networks and infectious disease: The colorado springs study. *Social science & medicine*, 38(1):79–88, 1994.

[141] Fredrik Liljeros, Christofer R. Edling, Luis A. Nunes Amaral, H. Eugene Stanley, and Yvonne Åberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[142] Peter Mariolis. Interlocking directorates and control of corporations: The theory of bank control. *Social Science Quarterly*, 56(3):425–439, 1975.

[143] Mark S. Mizruchi and David Bunting. Influence in corporate networks: An examination of four measures. *Administrative Science Quarterly*, 26(3):475–489, 1981.

[144] Guido Caldarelli. *Large scale structure and dynamics of complex networks: from information technology to finance and natural science*, volume 2. World Scientific, 2007.

[145] Ernesto Estrada and Philip A. Knight. *A first course in network theory*. Oxford University Press, USA, 2015.

[146] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97 – 125, 2012. Temporal Networks.

[147] Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences*, 101(40):14333–14337, 2004.

[148] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert-László Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98:158702, 2007.

[149] José L. Iribarren and Esteban Moro. Impact of human activity patterns on the dynamics of information diffusion. *Physical review letters*, 103(3):038702, 2009.

[150] Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber. Evidence for a bimodal distribution in human communication. *Proceedings of the national academy of sciences*, 107(44):18803–18808, 2010.

[151] Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th international conference on World Wide Web*, pages 915–924, 2008.

[152] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[153] Siemon de Lange, Marcel de Reus, and Martijn Van Den Heuvel. The laplacian spectrum of neural networks. *Frontiers in Computational Neuroscience*, 7:189, 2014.

[154] Daniel Korenblum. Laplacian mixture modeling for network analysis and unsupervised learning on graphs. *PLOS ONE*, 13(10):1–33, 2018.

[155] Ralucca Gera, Lázaro Alonso, Brian Crawford, Jeffrey House, J. A. Mendez-Bermudez, Thomas Knuth, and Ryan Miller. Identifying network structure similarity using spectral graph theory. *Applied network science*, 3(1):2, 2018.

[156] Jean-Charles Delvenne, Renaud Lambiotte, and Luis E. Rocha. Diffusion on networked systems is a question of time or structure. *Nature communications*, 6(1):1–10, 2015.

[157] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics Reports*, 716-717:1 – 58, 2017. Random walks and diffusion on networks.

[158] Florian Dörfler, Michael Chertkov, and Francesco Bullo. Synchronization in complex oscillator networks and smart grids. *Proceedings of the National Academy of Sciences*, 110(6):2005–2010, 2013.

[159] Melvyn Tyloo, Laurent Pagnier, and Philippe Jacquod. The key player problem in complex oscillator networks and electric power grids: Resistance centralities identify local vulnerabilities. *Science Advances*, 5(11), 2019.

[160] Meng Zhan, Shuai Liu, and Zhiwei He. Matching rules for collective behaviors on complex networks: Optimal configurations for vibration frequencies of networked harmonic oscillators. *PloS one*, 8(12):e82161, 2013.

[161] Melvyn Tyloo, Tommaso Coletta, and Philippe Jacquod. Robustness of synchrony in complex networks and generalized kirchhoff indices. *Phys. Rev. Lett.*, 120:084101, 2018.

[162] Adolf Fick. Über Diffusion. *Annalen der Physik*, 170(1):59–86, 1855.

[163] Douglas J. Klein and Milan Randić. Resistance distance. *Journal of mathematical chemistry*, 12(1):81–95, 1993.

[164] Roger Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955.

[165] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.

[166] Chao Yang, Robert Harkreader, Jialong Zhang, Seungwon Shin, and Guofei Gu. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 71–80, New York, NY, USA, 2012.

[167] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[168] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[169] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, 101(9):2658–2663, 2004.

[170] Mark E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[171] Mark E. J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69:066133, 2004.

[172] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Phys. Rev. E*, 72:027104, 2005.

[173] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[174] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific American*, 288(5):60–69, 2003.

[175] Paul Erdös and Alfréd Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[176] Réka Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 118(21):4947–4957, 2005.

[177] Eivind Almaas, Alexei Vazquez, and Albert-Laszlo Barabasi. Scale-free networks in biology. *Biological networks*, 3, 01 2013.

[178] Mark E.J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[179] Xavier Gabaix. Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 1999.

[180] George Kingsley Zipf. Human behaviour and the principle of least effort, adisson. *Wesley Press, Cambridge*, 6(3):306, 1949.

[181] Rajeev Kohli and Raaj Kumar Sah. Market shares: Some power law results and observations. *Management Science*, 52(11):1792–1798, 2006.

[182] William J. Reed and Barry D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys. Rev. E*, 66:067103, 2002.

[183] Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[184] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[185] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[186] Tiago P. Peixoto. Entropy of stochastic blockmodel ensembles. *Phys. Rev. E*, 85:056122, 2012.

[187] Tom A. Snijders and Krzysztof Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of classification*, 14(1):75–100, 1997.

[188] Antoine Moinet, Michele Starnini, and Romualdo Pastor-Satorras. Random walks in non-poissoinan activity driven temporal networks. *New Journal of Physics*, 21(9):093032, 2019.

[189] Laura Alessandretti, Kaiyuan Sun, Andrea Baronchelli, and Nicola Perra. Random walks on activity-driven networks with attractiveness. *Phys. Rev. E*, 95:052318, 2017.

[190] Nicola Perra, Andrea Baronchelli, Delia Mocanu, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Random walks and search in time-varying networks. *Phys. Rev. Lett.*, 109:238701, 2012.

[191] Matthieu Nadini, Kaiyuan Sun, Enrico Ubaldi, Michele Starnini, Alessandro Rizzo, and Nicola Perra. Epidemic spreading in modular time-varying networks. *Scientific reports*, 8(1):1–11, 2018.

[192] Iacopo Pozzana, Kaiyuan Sun, and Nicola Perra. Epidemic spreading on activity-driven networks with attractiveness. *Phys. Rev. E*, 96:042310, 2017.

[193] Ping Hu, Li Ding, and Xuming An. Epidemic spreading with awareness diffusion on activity-driven networks. *Phys. Rev. E*, 98:062322, 2018.

[194] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling contagion processes in activity driven networks. *Phys. Rev. Lett.*, 112:118702, 2014.

[195] Mingwu Li and Harry Dankowicz. Impact of temporal network structures on the speed of consensus formation in opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 523:1355 – 1370, 2019.

[196] Lorenzo Zino, Alessandro Rizzo, and Maurizio Porfiri. Consensus over activity-driven networks. *IEEE Transactions on Control of Network Systems*, 7(2):866–877, 2020.

[197] Thilo Gross and Bernd Blasius. Adaptive coevolutionary networks: a review. *Journal of The Royal Society Interface*, 5(20):259–271, 2008.

[198] Solomon E. Asch. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70(9):1, 1956.

[199] Robert P. Abelson. Mathematical models in social psychology. *Advances in experimental social psychology*, 3:1–54, 1967.

[200] Anton V. Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, 43:65–79, 2017.

[201] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[202] Michael Taylor. Towards a mathematical theory of influence and attitude change. *Human Relations*, 21(2):121–139, 1968.

[203] Wei Ren and Randal W. Beard. *Distributed consensus in multi-vehicle cooperative control*, volume 27. Springer, 2008.

[204] Andrea Baronchelli, Maddalena Felici, Vittorio Loreto, Emanuele Caglioti, and Luc Steels. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06), 2006.

[205] Sara Brin Rosenthal, Colin R Twomey, Andrew T Hartnett, Hai Shan Wu, and Iain D Couzin. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112(15):4690–4695, 2015.

[206] Yael Katz, Kolbjørn Tunstrøm, Christos C. Ioannou, Cristián Huepe, and Iain D. Couzin. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, 108(46):18720–18725, 2011.

[207] Herbert G. Tanner, Ali Jadbabaie, and George J. Pappas. Stable flocking of mobile agents, part i: Fixed topology. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 2, pages 2010–2015. IEEE, 2003.

[208] Vijay Gupta, Babak Hassibi, and Richard M. Murray. Stability analysis of stochastically varying formations of dynamic agents. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 1, pages 504–509. IEEE, 2003.

[209] Alejandro Rodriguez-Angeles and Henk Nijmeijer. Cooperative synchronization of robots via estimated state feedback. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, volume 2, pages 1514–1519. IEEE, 2003.

[210] Justin Werfel, Kirstin Petersen, and Radhika Nagpal. Designing collective behavior in a termite-inspired robot construction team. *Science*, 343(6172):754–758, 2014.

[211] Iain Couzin. Collective minds. *Nature*, 445(7129):715–715, 2007.

[212] Wei Ren, Randal W. Beard, and Ella M. Atkins. Information consensus in multivehicle cooperative control. *IEEE Control systems magazine*, 27(2):71–82, 2007.

[213] Hai-Tao Zhang, Zhaomeng Cheng, Guanrong Chen, and Chunguang Li. Model predictive flocking control for second-order multi-agent systems with input constraints. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 62(6):1599–1606, 2015.

[214] Wei Ren, Randal W. Beard, and Ella M. Atkins. A survey of consensus problems in multi-agent coordination. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 1859–1864. IEEE, 2005.

[215] Wei Ren and Ella Atkins. Second-order consensus protocols in multiple vehicle systems with local interactions. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, page 6238, 2005.

[216] Wei Ren and Ella Atkins. Distributed multi-vehicle coordinated control via local information exchange. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 17(10-11):1002–1033, 2007.

[217] Wen Yang, Andrea L. Bertozzi, and Xiaofan Wang. Stability of a second order consensus algorithm with time delay. In *2008 47th IEEE Conference on Decision and Control*, pages 2926–2931. IEEE, 2008.

[218] Jiahu Qin, Changbin Yu, and Sandra Hirche. Stationary consensus of asynchronous discrete-time second-order multi-agent systems under switching topology. *IEEE Transactions on Industrial Informatics*, 8(4):986–994, 2012.

[219] Wei Ren. Second-order consensus algorithm with extensions to switching topologies and reference models. In *2007 American Control Conference*, pages 1431–1436. IEEE, 2007.

[220] Wenwu Yu, Guanrong Chen, Ming Cao, and Jürgen Kurths. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(3):881–891, 2009.

[221] Wenwu Yu, Guanrong Chen, and Ming Cao. Some necessary and sufficient conditions for second-order consensus in multi-agent dynamical systems. *Automatica*, 46(6): 1089–1095, 2010.

[222] Darina Goldin and Jörg Raisch. Consensus for agents with double integrator dynamics in heterogeneous networks. *Asian Journal of Control*, 16(1):30–39, 2014.

[223] Xiaodong Ai, Shiji Song, and Keyou You. Second-order consensus of multi-agent systems under limited interaction ranges. *Automatica*, 68:329–333, 2016.

[224] Jiangping Hu, Guanrong Chen, and Han-Xiong Li. Distributed event-triggered tracking control of second-order leader-follower multi-agent systems. In *Proceedings of the 30th Chinese control conference*, pages 4819–4824. IEEE, 2011.

[225] Timur Kuran. The inevitability of future revolutionary surprises. *American Journal of Sociology*, 100(6):1528–1551, 1995.

[226] Rosabeth Moss Kanter. Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82(5):965–990, 1977.

[227] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. Experimental evidence for tipping points in social convention. *Science*, 360(6393):1116–1119, 2018.

[228] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*, pages 273–274, 2016.

[229] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature communications*, 9(1):1–9, 2018.

[230] Jacob Ratkiewicz, Michael D. Conover, Bruno Goncalves, Filippo Menczer, Alessandro Flammini, and Mark Meiss. Detecting and Tracking Political Abuse in Social Media. In *Proc. 5th of Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, July 2011.

[231] Mauro Mobilia, Alexander Petersen, and Sid Redner. On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08029–P08029, 2007.

[232] Mauro Mobilia. Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.*, 91:028701, Jul 2003.

[233] Pascal P. Klamser, Marc Wiedermann, Jonathan F. Donges, and Reik V. Donner. Zealotry effects on opinion dynamics in the adaptive voter model. *Phys. Rev. E*, 96:052315, 2017.

[234] Ercan Yildiz, Asuman Ozdaglar, Daron Acemoglu, Amin Saberi, and Anna Scaglione. Binary opinion dynamics with stubborn agents. *ACM Transactions on Economics and Computation (TEAC)*, 1(4):1–30, 2013.

[235] Naoki Masuda. Opinion control in complex networks. *New Journal of Physics*, 17(3): 033031, 2015.

[236] Gunjan Verma, Ananthram Swami, and Kevin Chan. The impact of competing zealots on opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 395:310 – 331, 2014.

[237] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. K. Szymanski. Social consensus through the influence of committed minorities. *Phys. Rev. E*, 84:011130, 2011.

[238] Alex Waagen, Gunjan Verma, Kevin Chan, Ananthram Swami, and Raissa D'Souza. Effect of zealotry in high-dimensional opinion dynamics models. *Phys. Rev. E*, 91:022811, 2015.

[239] Gert Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.

[240] Stacy Patterson, Yuhao Yi, and Zhongzhi Zhang. A resistance-distance-based approach for optimal leader selection in noisy consensus networks. *IEEE Transactions on Control of Network Systems*, 6(1):191–201, 2019.

[241] Prabir Barooah and Joao Hespanha. Graph effective resistance and distributed control: Spectral properties and applications. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3479–3485, 2006.

[242] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, page 57–66, New York, NY, USA, 2001.

[243] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In Xiaotie Deng and Fan Chung Graham, editors, *Internet and Network Economics*, pages 306–311, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[244] Ryan Rossi and Nesreen Ahmed. The network data repository with interactive graph analytics and visualization. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[245] Mehdi Alighanbari and Jonathan P. How. Decentralized task assignment for unmanned aerial vehicles. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 5668–5673, 2005.

[246] Kevin L. Moore and Dennis Lucarelli. Decentralized adaptive scheduling using consensus variables. *International Journal of Robust and Nonlinear Control*, 17(10-11):921–940, 2007.

[247] Sansan Li, Na Sun, Li Chen, and Xingang Wang. Network synchronization with periodic coupling. *Phys. Rev. E*, 98:012304, 2018.

[248] Sansan Li, Xingang Wang, and Shuguang Guan. Periodic coupling suppresses synchronization in coupled phase oscillators. *New Journal of Physics*, 20(11):113013, 2018.

[249] Theodore W. Grunberg and Dennice F. Gayme. Performance measures for linear oscillator networks over arbitrary graphs. *IEEE Transactions on Control of Network Systems*, 5 (1):456–468, March 2018.

[250] Francisco A. Rodrigues, Thomas K. DM. Peron, Peng Ji, and Jürgen Kurths. The kuramoto model in complex networks. *Physics Reports*, 610:1 – 98, 2016.

[251] Lawrence Turyn. The damped mathieu equation. *Quarterly of applied mathematics*, 51 (2):389–398, 1993.

[252] Lev D. Landau and Evgeny M. Lifschitz. Lehrbuch der Theoretischen Physik. Band 1: Mechanik, 1984.

[253] George William Hill. On the part of the motion of the lunar perigee which is a function of the mean motions of the sun and moon. *Acta mathematica*, 8(1):1–36, 1886.

[254] Andries E. Brouwer and Willem H. Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.

[255] Homero Gil de Zúñiga and Sebastián Valenzuela. The mediating path to a stronger citizenship: Online and offline networks, weak ties, and civic engagement. *Communication Research*, 38(3):397–421, 2011.

[256] Three and a half degrees of separation. https://research.fb.com/blog/2016/02/three-and-a-half-degrees-of-separation/. Accessed: 2020-01-05.

[257] Christopher P. Earley and Elaine Mosakowski. Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, 43(1): 26–49, 2000.

[258] Edward L. Glaeser and Bryce A. Ward. Myths and realities of american political geography. *The Journal of Economic Perspectives*, 20(2):119–144, 2006.

[259] Aaron M. McCright and Riley E. Dunlap. The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *The Sociological Quarterly*, 52(2):155–194, 2011.

[260] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

[261] Michael D. Conover, Jacob Ratkiewicz, Matthew R. Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. *Icwsm*, 133(26): 89–96, 2011.

[262] Alexander Hanna, Chris Wells, Peter Maurer, Lew Friedland, Dhavan Shah, and Jörg Matthes. Partisan alignments and political polarization online: A computational approach to understanding the french and us presidential elections. In *Proceedings of the 2nd Workshop on Politics, Elections and Data*, PLEAD '13, page 15–22, New York, NY, USA, 2013.

[263] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. Falling into the echo chamber: the italian vaccination debate on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 130–140, 2020.

[264] Guillaume Deffuant, David Neau, Frédéric Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, (3):11, 2001.

[265] Rainer Hegselmann and Ulrich Krause. Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation*, 5 (3):1–2, 2002.

[266] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007.

[267] Takasumi Kurahashi-Nakamura, Michael Mäs, and Jan Lorenz. Robust clustering in generalized bounded confidence models. *Journal of Artificial Societies and Social Simulation*, 19(4):7, 2016.

[268] Miguel Pineda, Raul Toral, and Emilio Hernández-García. Noisy continuous-opinion dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(08), 2009.

[269] Wander Jager and Frédéric Amblard. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Comput. Math. Organ. Theory*, 10(4):295–303, January 2005.

[270] Noah Mark. Culture and competition: Homophily and distancing explanations for cultural niches. *American Sociological Review*, 68:319–345, 2003.

[271] Milton E Rosenbaum. The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology*, 51(6):1156, 1986.

[272] Thomas Feliciani, Andreas Flache, and Jochem Tolsma. How, when and where can spatial segregation induce opinion polarization? two competing models. *Journal of Artificial Societies and Social Simulation*, 20(2):6, 2017.

[273] Károly Takács, Andreas Flache, and Michael Mäs. Discrepancy and disliking do not induce negative opinion shifts. *PLOS ONE*, 11(6):1–21, 2016.

[274] Jae K. Shin. Information accumulation system by inheritance and diffusion. *Physica A: Statistical Mechanics and its Applications*, 388(17):3593 – 3599, 2009.

[275] Kristina Lerman and Rumi Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), 2010.

[276] Sebastian Stier, Johannes Breuer, Pascal Siegers, and Kjerstin Thorson. Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5):503–516, 2020.

[277] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. Evidence of complex contagion of information in social media: An experiment using twitter bots. *PLOS ONE*, 12(9):1–12, 2017.

[278] Lada A. Adamic, Thomas M. Lento, Eytan Adar, and Pauline C. Ng. Information evolution in social networks. In *Proceedings of the ninth ACM international conference on web search and data mining*, pages 473–482, 2016.

[279] Kar Wai Lim and Wray Buntine. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 1319–1328, 2014.

[280] Diana Maynard and Adam Funk. Automatic detection of political opinions in tweets. In Raúl García-Castro, Dieter Fensel, and Grigoris Antoniou, editors, *The Semantic Web: ESWC 2011 Workshops*, pages 88–99, Berlin, Heidelberg, 2012.

[281] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.

[282] Thorsten Faas. Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, (52):120–135, 2003.

[283] Agust Arnorsson and Gylfi Zoega. On the causes of brexit. *European Journal of Political Economy*, 55:301 – 323, 2018.

[284] Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. Content and network dynamics behind egyptian political polarization on twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 2015, Vancouver, BC, Canada, March 14 - 18, 2015*, pages 700–711, 2015.

[285] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*, 6:37825, 2016.

[286] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. Science vs conspiracy: Collective narratives in the age of misinformation. *PLOS ONE*, 10(2):1–17, 02 2015.

[287] T. Vaz Martins, Miguel Pineda, and Raul Toral. Mass media and repulsive interactions in continuous-opinion dynamics. *EPL (Europhysics Letters)*, 91(4):48003, 2010.

[288] Heather Z. Brooks and Mason A. Porter. A model for the influence of media on the ideology of content in online social networks. *Phys. Rev. Research*, 2:023041, 2020.

[289] Tingyu Li and Hengmin Zhu. Effect of the media on the opinion dynamics in online social networks. *Physica A: Statistical Mechanics and its Applications*, 551:124117, 2020.

[290] Rainer Hegselmann and Ulrich Krause. Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Networks & Heterogeneous Media*, 10(3):477, 2015.

[291] Bertrand Jayles, Hye-rin Kim, Ramón Escobedo, Stéphane Cezera, Adrien Blanchet, Tatsuya Kameda, Clément Sire, and Guy Theraulaz. How social information can improve estimation accuracy in human groups. *Proceedings of the National Academy of Sciences*, 114(47):12620–12625, 2017.

[292] Gordon Brown and Nick Chater. The chronological organization of memory: Common psychological foundations for remembering and timing. *Time and Memory: Issues in Philosophy and Psychology*, pages 77–110, 2001.

[293] Kathleen Carley. A theory of group stability. *American Journal of Sociology*, 56(3): 331–354, 1991.

[294] H. Sompolinsky, A. Crisanti, and H. J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262, Jul 1988. doi: 10.1103/PhysRevLett.61.259. URL https://link.aps.org/doi/10.1103/PhysRevLett.61.259.

[295] Eran Eldar, Jonathan D Cohen, and Yael Niv. The effects of neural gain on attention and learning. *Nature neuroscience*, 16(8):1146–1153, 2013.

[296] Simon Schweighofer, Frank Schweitzer, and David Garcia. A weighted balance model of opinion hyperpolarization. *Journal of Artificial Societies and Social Simulation*, 23(3):5, 2020.

[297] Michele Starnini and Romualdo Pastor-Satorras. Topological properties of a time-integrated activity-driven network. *Phys. Rev. E*, 87:062807, 2013.

[298] Antoine Moinet, Michele Starnini, and Romualdo Pastor-Satorras. Burstiness and aging in social temporal networks. *Phys. Rev. Lett.*, 114:108701, 2015.

[299] Magdalini Eirinaki, Jerry Gao, Iraklis Varlamis, and Konstantinos Tserpes. Recommender systems for large-scale social networks: A review of challenges and solutions. *Future Generation Computer Systems*, 78:413 – 418, 2018.

[300] Jon A. Krosnick. Attitude importance and attitude change. *Journal of Experimental Social Psychology*, 24(3):240 – 255, 1988.

[301] Venkata Rama Kiran Garimella and Ingmar Weber. A long-term analysis of polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 528–531, 2017.

[302] Haokai Lu, James Caverlee, and Wei Niu. Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 213–222, 2015.

[303] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. Users polarization on facebook and youtube. *PLOS ONE*, 11(8):1–24, 2016.

[304] Michele Starnini, Mattia Frasca, and Andrea Baronchelli. Emergence of metapopulations and echo chambers in mobile agents. *Scientific reports*, 6:31834, 2016.

[305] Christopher Currin and Ali Khaledi-Nasab. Depolarization of echo chambers by random dynamical nudge, 2021. URL https://arxiv.org/abs/2101.04079. arXiv preprint.

[306] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[307] Martin Hilbert and Priscila López. The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025):60–65, 2011.

[308] Rie Kubota Ando. Latent semantic space: Iterative scaling improves precision of inter-document similarity measurement. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, page 216–223, 2000.

[309] Yphtach Lelkes. Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, 80(S1):392–410, 2016.

[310] Andreas Flache and Michael Mäs. Why do faultlines matter? a computational model of how strong demographic faultlines undermine team cohesion. *Simulation Modelling Practice and Theory*, 16(2):175 – 191, 2008.

[311] Eric Molleman. Diversity in demographic characteristics, abilities and personality traits: Do faultlines affect team functioning? *Group decision and Negotiation*, 14(3):173–193, 2005.

[312] Jeffrey Pfeffer. Organizational demography: Implications for management. *California Management Review*, 28(1):67–81, 1985.

[313] Donn Byrne, Gerald L. Clore, and George Smeaton. The attraction hypothesis: Do similar attitudes affect anything? *Journal of Personality and Social Psychology*, 51(6): 1167–1170, 1986.

[314] Fang Fang Chen and Douglas T Kenrick. Repulsion or attraction? Group membership and assumed attitude similarity. *Journal of personality and social psychology*, 83(1):111, 2002.

[315] Thomas K. Landauer, Danielle S. McNamara, Simon Dennis, and Walter Kintsch. *Handbook of latent semantic analysis*. Psychology Press, 2013.

[316] The rise of social media. https://ourworldindata.org/rise-of-social-media. Accessed: 2021-01-25.

[317] David Westerman, Patric R. Spence, and Brandon Van Der Heide. Social media as information source: Recency of updates and credibility of information. *Journal of computer-mediated communication*, 19(2):171–183, 2014.

[318] Hitkul, Avinash Prabhu, Dipanwita Guhathakurta, Jivitesh jain, Mallika Subramanian, et al. Capitol (pat)riots: A comparative study of twitter and parler, 2021. URL https://arxiv.org/abs/2101.06914. arXiv preprint.

[319] Philipp Lorenz-Spreen, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature human behaviour*, 4(11):1102–1109, 2020.

[320] Ralph Hertwig and Till Grüne-Yanoff. Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6):973–986, 2017.

[321] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D. Wherry, and Natalya N. Bazarova. Upstanding by design: Bystander intervention in cyberbullying. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–12, 2018.

[322] Bailey K. Fosdick, Daniel B. Larremore, Joel Nishimura, and Johan Ugander. Configuring random graph models with fixed degree sequences. *SIAM Review*, 60(2):315–355, 2018.

[323] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007.